stochastic
processes
and their
applications

ELSEVIER Stochastic Processes and their Applications 91 (2001) 205–238
www.elsevier.com/locate/spa

Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems

Liming Wu

Laboratoire de Mathématiques Appliquées, CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubiere, France
Department of Mathematics, Wuhan University, People's Republic of China

Received 1 April 1999; received in revised form 1 February 2000; accepted 1 July 2000

Abstract

A classical damping Hamiltonian system perturbed by a random force is considered. The locally uniform large deviation principle of Donsker and Varadhan is established for its occupation empirical measures for large time, under the condition, roughly speaking, that the force driven by the potential grows infinitely at infinity. Under the weaker condition that this force remains greater than some positive constant at infinity, we show that the system converges to its equilibrium measure with exponential rate, and obeys moreover the moderate deviation principle. Those results are obtained by constructing appropriate Lyapunov test functions, and are based on some results about large and moderate deviations and exponential convergence for general strong-Feller Markov processes. Moreover, these conditions on the potential are shown to be sharp. © 2001 Elsevier Science B.V. All rights reserved.

MSC: 60F10; 93E15; 60H10; 70L05

Keywords: Stochastic Hamiltonian systems; Large deviations; Moderate deviations; Exponential convergence; Hyper-exponential recurrence

0. Introduction

Let us consider a classical damping Hamiltonian system, perturbed by a random force. More precisely, let x_t (resp. y_t) be the position (resp. the velocity) at time $t \geqslant 0$, of a physical system moving in $\mathbb{R}^d$, under the action of the three forces:

(1) the force $-\nabla V(x_t)$ driven by the potential V;
(2) the damping force $-c(x_t, y_t)y_t$, where $c(x,y) = (c_{ij}(x,y))_{1 \leqslant i,j \leqslant d}$ is the damping coefficient;
(3) the random force modeled as $\Sigma(x_t, y_t)(\mathrm{d}W_t/\mathrm{d}t)$, where (W_t) is a standard Brownian Motion in $\mathbb{R}^d$ and $\Sigma(x,y) = (\Sigma_{ij}(x,y))_{1 \leqslant i,j \leqslant d}$ describes the strength of the random perturbation.

E-mail address: wuliming@ucfma.univ-bpclermont.fr (L. Wu).

0304-4149/01/$ - see front matter
PII: S0304-4149(00)00061-2

Hence $(Z_t := (x_t, y_t) \in \mathbb{R}^{2d}, t \geqslant 0)$ is governed by the following Ito stochastic differential equation (in short: s.d.e.):

$$\begin{aligned} \mathrm{d}x_t &= y_t\,\mathrm{d}t \\ \mathrm{d}y_t &= \Sigma(x_t, y_t)\mathrm{d}W_t - (c(x_t, y_t)y_t + \nabla V(x_t))\,\mathrm{d}t. \end{aligned} \tag{0.1}$$

Throughout this paper, for diffusion (0.1) we assume that

(H1) the potential V is lower bounded and continuously differentiable over $\mathbb{R}^d$;

(H2) the damping coefficient $c(x, y)$ is continuous and for all $N > 0$: $\sup_{|x|\leqslant N, y\in\mathbb{R}^d} \|c(x, y)\|_{\mathrm{H.S.}} < +\infty$, and there exist $c, L > 0$ so that $c^s(x, y) \geqslant cI > 0, \forall(|x| > L, y \in \mathbb{R}^d)$;

(H3) the random strength Σ is symmetric, infinitely differentiable and for some $\sigma > 0$: $0 < \Sigma(x, y) \leqslant \sigma I$ over $\mathbb{R}^{2d}$.

Here $c^s(x, y)$ is the symmetrization of the matrix $c(x, y)$, given by $(\frac{1}{2}(c_{ij}(x, y) + c_{ji}(x, y))_{1\leqslant i,j\leqslant d}$, $\|\cdot\|_{\mathrm{H.S.}}$ is the Hilbert–Schmidt norm of matrix; the order relation on symmetric matrices is the usual one defined by the definite non-negativeness; and "$\Sigma > 0$" means that it is strictly positive definite.

Let us first consider the particular but current situation where $c(x, y) \equiv cI > 0$ and $\Sigma(x, y) \equiv \sigma I > 0$.

When $\sigma = 0$ (no random perturbation), the system is *dissipative* because of the existence of the damping force $-cy$, and it will converge to the phase points where the Hamiltonian $H(x, y) = \frac{1}{2}|y|^2 + V(x)$ attains the local minima.

When $\sigma > 0$, the random force will compensate the loss of energy caused by the damping force, and the system will approach some non-degenerate equilibrium measure. Indeed in this particular case, (0.1) has a unique invariant measure (up to a numerical constant factor), given by

$$\alpha(\mathrm{d}x, \mathrm{d}y) = \exp\left(-\frac{2c}{\sigma^2}H(x, y)\right)\mathrm{d}x\,\mathrm{d}y \tag{0.2}$$

where $H(x, y)$ is the Hamiltonian (see, e.g. Roberts and Spanos, 1990).

The asymptotic behavior of (0.1) with $c(x, y) \equiv cI$ is widely studied both in the cases $c = 0$ or $c > 0$. When $c = 0$, (0.1) becomes the so called stochastic gradient Hamiltonian system, see the works of Albeverio and Klar (1994), Albeverio and Kolokoltsov (1997), Freidlin and Weber (1998) and the references therein. Those studies are mainly devoted to long-time behavior of the system: the transience, the scattering theory and the averaging principle, etc. A general remark: the only invariant measure of the system is the Liouville measure $\mathrm{d}x\,\mathrm{d}y$, which is infinite. Hence none of the usual ergodic properties, such as positive recurrence, large deviations or moderate deviations etc, holds.

The situation where $c > 0$ and α is finite is interesting at least from two points of view:

(1) it models many random vibration phenomena, see Arnold (1974) and Roberts and Spanos (1990);

(2) since the marginal law of the equilibrium measure α in x (resp. y) is the Gibbs measure $\exp(-(2c/\sigma^2)V(x))\,\mathrm{d}x$ (resp. the Gaussian measure $\exp(-(c/\sigma^2)|y|^2)$),

α describes exactly the equilibrium statistical mechanical state, and (0.1) can be employed to model the microscopical behavior of N-particles system ($d=3N$ very large). This insight observation goes back up to Langevin.

See the book of Khas'minskii (1980) for studies on positive recurrence and ergodic properties.

The model (0.1) is quite general, for example it covers the generalized Duffing oscillator ($c(x,y)=c>0$ and $V(x)$ is a lower bounded polynomial), the van der Pol oscillator ($c(x,y)=x^2-1$, $V(x)=\frac{1}{2}\omega_0^2x^2$), etc.

The main aim of this paper is to study the exponential convergence of (0.1), the large deviation principle (in short: LDP) of Donsker and Varadhan and the moderate deviation principle (in short: MDP) for the occupation empirical measures

$$L_t := \frac{1}{t}\int_0^t \delta_{Z_s}\,\mathrm{d}s, \quad Z_s := (x_s, y_s) \tag{0.3}$$

(where $\delta_\cdot$ denotes the Dirac measure), and for the process-level empirical measures

$$R_t := \frac{1}{t}\int_0^t \delta_{Z_{s+\cdot}}\,\mathrm{d}s \tag{0.4}$$

as t goes to infinity. Here $Z_{s+\cdot}$ denotes the path $t \to Z_{s+t}$, a random element in $C(\mathbb{R}^+, \mathbb{R}^{2d})$.

Roughly speaking, in order to get those three strong ergodic properties, the force $-\nabla V(x)$ should be strong enough for $|x|$ large, to make the system return quickly to the compact subsets of $\mathbb{R}^{2d}$. A quite natural condition for this intuitive picture is

$$\nabla V(x)\cdot x/|x| \to +\infty \quad \text{as } |x| \to +\infty \tag{0.5}$$

or

$$\liminf_{|x|\to+\infty} \nabla V(x)\cdot x/|x| > 0. \tag{0.6}$$

Condition (0.5) (resp. (0.6)) means that the component of the force $-\nabla V$ in the direction to the origin grows infinite (resp. remains greater than some positive constant) as $|x|$ goes to infinity. Our aim is to explore several consequences of (0.5) and (0.6) in the asymptotic behavior of the system (0.1).

This paper is organized as follows. In the preliminary Section 1, we first show the existence and the uniqueness of the weak solution of (0.1) and a Girsanov formula. Next we prove the strong Feller property, which is basic for all results in this paper.

In Section 2, we discuss general strong Feller Markov processes. A necessary and sufficient condition both for the level-2 and level-3 LDP of Donsker and Varadhan is given in Theorem 2.1, by means of the hyper-exponential recurrence (a notion introduced here). As corollaries, we present a Lyapunov test function type criterion, originated from the pioneering works of Donsker and Varadhan (1975, 1976, 1983), and we discuss also large deviations for unbounded additive functionals.

For the exponential convergence we first recall in Theorem 2.4 the criterion of Lyapunov test function due to Down et al. (1995). As its consequence (as in Wu (1995)) we obtain the MDP in Theorems 2.6 and 2.7. To keep the continuity of presentation, the proofs of Theorems 2.1, 2.6 and 2.7 are left to the Appendix.

As applications of those general results, we obtain the LDP of (0.1) under some more general conditions than (0.5) (in the multi-dimensional case) in Section 3, and the exponential convergence and the MDP under (0.6) in Section 4. The key for those results is construction of an appropriate Lyapunov test function.

In Section 5 we show at first that condition (0.5) and (0.6) are sharp for the LDP and for the exponential convergence, respectively. Applications to the generalized Duffing oscillator and to the van der Pol model are quickly examined.

1. Preliminaries

In this preliminary section we shall establish several basic facts about (0.1), such as the existence and the uniqueness of solution, the Girsanov formula and the strong Feller property.

1.1. Notations

We begin with some necessary notations.

The Euclidean inner product in $\mathbb{R}^{2d}$ or $\mathbb{R}^d$ is denoted by $\cdot$, and $|z| := \sqrt{z\cdot z}$. The Borel σ-field of $\mathbb{R}^{2d}$ is denoted by $\mathscr{B}$. As usual notation, $C^m(\mathbb{R}^d)$ (resp. $C_0^m(\mathbb{R}^d)$) denotes the space of all real m-times continuously differentiable functions (resp. and with compact support) on $\mathbb{R}^d$. Let $C^{m,n}(\mathbb{R}^{2d})$ (resp. $C_b^{m,n}(\mathbb{R}^{2d})$) be the space of all functions $f(x,y)$ such that $\partial_x^k f,\ k=0,1,\dots,m$ and $\partial_y^l f,\ l=1,\dots,n$ are continuous (resp. and bounded) on $\mathbb{R}^{2d}$. We write simply $C(\mathbb{R}^{2d})$, $C_b(\mathbb{R}^{2d})$ for $C^{0,0}(\mathbb{R}^{2d})$, $C_b^{0,0}(\mathbb{R}^{2d})$.

Given a σ-field $\mathscr{G}$, let $b\mathscr{G}$ be the space of all real bounded and $\mathscr{G}$-measurable functions.

Consider the space $\Omega := C(\mathbb{R}^+,\mathbb{R}^{2d})$ of continuous functions from $\mathbb{R}^+$ to $\mathbb{R}^{2d}$, equipped with the usual compact convergence topology. For $\omega\in\Omega$, let $Z_t(\omega) = \omega(t) = (x_t(\omega),\ y_t(\omega))$, $t\geqslant 0$ be the coordinates. Ω is equipped with the natural filtration $(\mathscr{F}_t := \sigma(Z_s,\ 0\leqslant s\leqslant t))_{t\geqslant 0}$.

The generator $\mathscr{L}$ of (0.1) is given by: for any $f\in C^{1,2}(\mathbb{R}^{2d})$,

$$\mathscr{L}f(x,y) := \frac{1}{2}\sum_{i,j=1}^{d}(\Sigma^2)_{ij}(x,y)\partial_{y_i}\partial_{y_j}f(x,y) + y\cdot\nabla_x f(x,y) - (c(x,y)y + \nabla_x V(x))\cdot\nabla_y f(x,y).$$

Recall that (H1), (H2) and (H3) for (0.1) are assumed throughout this paper (except explicit contrary statement). In this section we assume neither (0.5) nor (0.6).

1.2. A Girsanov formula

First of all we should show the existence and the uniqueness of the weak solution of (0.1). This is done in

Lemma 1.1. *For every initial state* $z=(x,y)\in\mathbb{R}^{2d}$, *the s.d.e.* (0.1) *admits a unique weak solution* $\boldsymbol{P}_z$ (*a probability measure on* Ω), *which is non-explosive. Moreover* $\boldsymbol{P}_z \ll \boldsymbol{P}_z^0$ *on* $(\Omega,\mathscr{F}_t)$ *for each* $t>0$, *and the Girsanov formula below holds*

$$\left.\frac{\mathrm{d}\boldsymbol{P}_z}{\mathrm{d}\boldsymbol{P}_z^0}\right|_{\mathscr{F}_t} = \exp\left(-\int_0^t \Sigma^{-1}(x_s,y_s)[c(x_s,y_s)y_s+\nabla_x V(x_s)]\mathrm{d}W_s \right.$$
$$\left. -\frac{1}{2}\int_0^t |\Sigma^{-1}(x_s,y_s)(c(x_s,y_s)y_s+\nabla_x V(x_s))|^2\,\mathrm{d}s\right), \tag{1.1}$$

where $\boldsymbol{P}_z^0$ *is the law of the solution of* (0.1) *associated with* $c(x,y)=0$ *and* $V=0$, *and* $(W_t := \int_0^t \Sigma^{-1}(x_s,y_s)\,\mathrm{d}y_s,\ t\geqslant 0)$ *is a standard Wiener process under* $\boldsymbol{P}_z^0$.

Proof. Recall at first that for $c(x,y)=0$ and $V=0$, the s.d.e. (0.1) has a unique strong solution which is non-explosive, by (H3).

Let $\boldsymbol{P}_z$ be a weak solution of (0.1) with life time (or explosion time) $\zeta = \sup_N \inf\{t\geqslant 0;\ |Z_t|\geqslant N\}$.

Let $\tau_R := \inf\{t\geqslant 0;\ |y_t|=R\}$ where $R>|y_0|=|y|$. Since $|x_t|\leqslant |x|+Rt$, $\forall t\leqslant \tau_R$, then $\sup_{R>0}\tau_R\leqslant\zeta$. By following the proof of (Jacod and Shiryaev, 1987, pp. 188–189, Theorem 5.38), we can show that the weak solution until $t\leqslant\tau_R$ of (0.1) is *unique* and it is given by

$$\boldsymbol{P}_z|_{\mathscr{F}_{\tau_R\wedge t}} = M_{t\wedge\tau_R}\cdot \mathrm{d}\boldsymbol{P}_z^0 \quad \forall t\geqslant 0 \tag{1.2}$$

where $(M_t)_{t\geqslant 0}$ is the exponential local martingale in the right-hand side (RHS in short) of (1.1) (Note: since $\Sigma(x_s,y_s)\geqslant a(R,t)I$ for all $0\leqslant s\leqslant t\wedge\tau_R$ where $a(R,t)$ is the infimum of the lowest eigenvalue of $\Sigma(\tilde{x},\tilde{y})$ for $(|\tilde{x}|\leqslant|x|+Rt, |\tilde{y}|\leqslant R)$, which is strictly positive by (H3), $M_{\cdot\wedge\tau_R}$ is then a true martingale by Novikov's criterion).

We shall show that for each $t>0$ fixed,

$$\lim_{R\to+\infty}\boldsymbol{P}_z(\tau_R>t)=1. \tag{1.3}$$

It implies not only the non-explosion of (0.1) (obvious), but also

$$\int M_t\,\mathrm{d}\boldsymbol{P}_z^0 \geqslant \int 1_{[\tau_R>t]}M_{t\wedge\tau_R}\,\mathrm{d}\boldsymbol{P}_z^0 = \boldsymbol{P}_z(\tau_R>t)\to 1 \quad \text{as } R\to\infty.$$

Hence (M_t) is a martingale. This last fact, combined with (1.2) and (1.3), implies the Girsanov formula (1.1), then the global uniqueness, too.

To show (1.3), let us consider a natural test function of Lyapunov type: the Hamiltonian $H(x,y)=\frac{1}{2}|y|^2+V(x)$. Letting $\mathrm{tr}(\cdot)$ be the trace of matrix $\cdot$, we have

$$\begin{aligned}\mathscr{L}H &= \frac{\mathrm{tr}\,\Sigma^2(x,y)}{2} + y\cdot\nabla_x V(x) - (c(x,y)y+\nabla_x V(x))\cdot y\\ &= \frac{\mathrm{tr}\,\Sigma^2(x,y)}{2} - c^s(x,y)y\cdot y.\end{aligned}$$

Since $c^s(x,y)\geqslant -AI$ over $\mathbb{R}^{2d}$ for some constant $A>0$ by (H2), and $\mathrm{tr}\,\Sigma^2(x,y)\leqslant \mathrm{d}\sigma^2$ by (H3), then $\tilde{H} := H - \inf_{\mathbb{R}^d} V + 1$ satisfies

$$\mathscr{L}\tilde{H}\leqslant b\tilde{H}$$

where $b := \max\{2A, d\sigma^2/2\}$. Consequently $(e^{-bt\wedge\tau_R}\tilde{H}(Z_{t\wedge\tau_R}))$ is a $\boldsymbol{P}_z$-supermartingale by Ito's formula. Noting that $\tilde{H}(Z_{\tau_R}) \geqslant R^2/2$ on $[\tau_R < +\infty]$, we get

$$\boldsymbol{P}_z(\tau_R \leqslant t) \leqslant \frac{2e^{bt}}{R^2}\boldsymbol{E}^{\boldsymbol{P}_z} 1_{[\tau_R\leqslant t]} e^{-bt\wedge\tau_R}\tilde{H}(Z_{t\wedge\tau_R}) \leqslant \frac{2e^{bt}}{R^2}\tilde{H}(z),$$

where (1.3) follows. □

1.3. Strong Feller property

The strong Feller property of (P_t) below is basic for all results in this paper:

Proposition 1.2. *Let $(P_t(z, dz'))_{t\geqslant 0}$ be the semigroup of transition probability kernels of the Markov process $((Z_t)_{t\geqslant 0}, (\boldsymbol{P}_z)_{z\in\mathbb{R}^{2d}})$ (solution of (0.1)). For every $t > 0$ and $z \in \mathbb{R}^{2d}$, $P_t(z, dz') = p_t(z, z')dz'$, $p_t(z, z') > 0$, dz'-a.e. and*

$$z \to p_t(z, \cdot) \text{ is continuous from } \mathbb{R}^{2d} \text{ to } L^1(\mathbb{R}^{2d}, dz'). \tag{1.4}$$

In particular P_t is strong Feller for each $t > 0$.

Remark 1.3. The main technical difficulty, as well known, comes from the degeneration of the diffusion (0.1). The strong Feller property above is far from being obvious, since the coefficients ∇V and $c(x, y)$ appeared in (0.1) are only continuous. If they are all C^∞, it is known that $P_t(z, dz') = p_t(z, z')dz'$, with $p_t \in C^\infty(\mathbb{R}^{2d}\times\mathbb{R}^{2d})$, by means of the hypoellipticity. This proposition improves a previous result due to Hilbert (1990).

Proof of Proposition 1.2. We begin with the case $c(x, y) = 0$ and $V = 0$. By (H3), the corresponding diffusion is hypoelliptic. By the hypoellipticity of Hörmander, the transition probability $P_t^0(z, dz')$ of $\boldsymbol{P}_z^0$ satisfies $P_t^0(z, dz') = p_t^0(z, z')\,dz'$ with $p_t^0 \in C^\infty(\mathbb{R}^{2d}\times\mathbb{R}^{2d})$ and $P_t^0(z, \cdot)$ is equivalent to the Lebesgue measure for every $z \in \mathbb{R}^{2d}$ and $t > 0$. Combined with the fact that $z \to \boldsymbol{P}_z^0$ is continuous with respect to (in short: w.r.t.) the weak convergence of measures on Ω, the previous property implies that P_t^0 is strong Feller for $t > 0$.

By the Girsanov formula (1.1), $P_t(z, dz') = p_t(z, z')\,dz'$, $p_t(z, z') > 0$, dz'-*a.e.* for every $t > 0$ and $z \in \mathbb{R}^{2d}$.

The property (1.4) for P_t is exactly the so called *strong Feller property in the strict sense* in Revuz (1976, Definition 5.8, p. 34). As $P_t = P_{t/2}P_{t/2}$, by the result of Revuz (1976, Theorem 5.10, p. 35), it is enough to show the strong Feller property of $P_{t/2}$, i.e., $P_t f \in C_b(\mathbb{R}^{2d})$ for any $f \in b\mathscr{B}$ and $t > 0$.

Our proof will be direct and elementary, based on the Girsanov formula (1.1) in Lemma 1.1.

Let $(W_t)_{t\geqslant 0}$ be a $\mathbb{R}^d$-valued standard Brownian Motion defined on some filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \boldsymbol{P})$ where the filtration satisfies the usual condition. We denote the *strong* solution of (0.1) with $c(x, y) = 0$ and $V = 0$ and with initial condition $Z_0 = z$ by $(Z_t^0(z))_{t\geqslant 0}$, which is defined on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \boldsymbol{P})$.

Fix $t > 0$ and $f \in b\mathscr{B}$. We have by Lemma 1.1 that for all $z \in \mathbb{R}^{2d}$,

$$P_t f(z) = \mathbb{E}^{\boldsymbol{P}_z} f(x_t, y_t) = \mathbb{E}^{\boldsymbol{P}} f(Z_t^0(z)) \cdot M_t(z) \tag{1.5}$$

where $(M_t(z))$ is the exponential martingale given by the RHS of (1.1) but with (x_s, y_s) substituted by the strong solution $(Z_s^0(z))$ specified previously.

Let (z_n) be a sequence of points in $\boldsymbol{R}^{2d}$ tending to z. Then $Z^0(z_n) \to Z^0(z)$ uniformly on the bounded time intervals in probability $\boldsymbol{P}$. Thus $M_t(z_n) \to M_t(z)$ in $\boldsymbol{P}$-probability too by a well known property of stochastic integral. On the other hand, $M_t(z_n)$ are nonnegative and by Lemma 1.1,

$$\mathbb{E}^{\boldsymbol{P}} M_t(z_n) = 1 = \mathbb{E}^{\boldsymbol{P}} M_t(z). \tag{1.6}$$

Now by an ingenious well known lemma, we conclude that

$$M_t(z_n) \to M_t(z) \quad \text{in } L^1(\boldsymbol{P}). \tag{1.7}$$

We show now

$$f(Z_t^0(z_n)) \to f(Z_t^0(z)) \quad \text{in } \boldsymbol{P}\text{-probability}. \tag{1.8}$$

To this end, fix a probability measure μ on $\mathbb{R}^{2\mathrm{d}}$, equivalent to the Lebesgue measure $\mathrm{d}z'$. Let $q_t^0(z,z')$ be the density of $\boldsymbol{P}(Z_t^0(z) \in \mathrm{d}z') = P_t^0(z, \mathrm{d}z')$ w.r.t. $\mu(\mathrm{d}z')$. By the strong Feller property of P_t^0 recalled above, for any $g \in L^\infty(\mu)$,

$$\int g(z') q_t^0(z_n, z') \mu(\mathrm{d}z') \to \int g(z') q_t^0(z,z') \mu(\mathrm{d}z'),$$

i.e., $q_t^0(z_n, \cdot) \to q_t^0(z, \cdot)$ in $\sigma(L^1(\mu), L^\infty(\mu))$. Then the family $\{q_t^0(z_n, \cdot); n\}$ is uniformly integrable in $L^1(\mu)$ by Dunford–Pettis theorem. Hence for any $\varepsilon > 0$, there is $\delta > 0$ such that for all $A \in \mathscr{B}$,

$$\mu(A) < \delta \Rightarrow \int q_t^0(z_n, z') 1_{[z' \in A]} \mu(\mathrm{d}z') = \boldsymbol{P}\left(Z_t^0(z_n) \in A\right) < \varepsilon, \ \forall n$$
$$\text{and } \boldsymbol{P}(Z_t^0(z) \in A) < \varepsilon.$$

On the other hand, by Egorov's Lemma, we can find a compact subset D such that $\mu(D^c) < \delta$ and $f|_D$ is uniformly continuous, i.e., for any $\eta > 0$, there exists $\delta_2 > 0$ such that

$$z, z' \in D \quad \text{and} \quad |z - z'| < \delta_2 \Rightarrow |f(z) - f(z')| < \eta.$$

We get therefore,

$$\begin{aligned}
&\boldsymbol{P}(|f(Z_t^0(z_n)) - f(Z_t^0(z))| > \eta) \\
&\quad \leqslant \boldsymbol{P}(|Z_t^0(z_n) - Z_t^0(z)| \geqslant \delta_2) + \boldsymbol{P}(Z_t^0(z_n)) \notin D) + \boldsymbol{P}(Z_t^0(z) \notin D) \\
&\quad \leqslant \boldsymbol{P}(|Z_t^0(z_n) - Z_t^0(z)| \geqslant \delta_2) + 2\varepsilon.
\end{aligned}$$

Letting $n \to \infty$, we get (1.8), because $\varepsilon, \eta > 0$ are arbitrary.

By (1.7), (1.8) and (1.5), $P_t f(z_n) \to P_t f(z)$ as n goes to infinity, the desired result. □

Remark. Lemma 1.1 and then Proposition 1.2 still hold under (H1)+(H3) and "$c^s(x, y) \geqslant -AI$ for some $A \in \mathbb{R}$ and $c(x, y)$ is continuous over $\mathbb{R}^{2d}$", instead of (H2). This can be seen from the proofs above. However, we do not know whether $\sigma \in C^\infty$ could be weakened as $\sigma \in C^1$.

2. Several general results for strong Feller Markov processes

In this section we present several general results about large and moderate deviations of general strong Feller Markov processes, which will allow us in the next sections to confine our studies to the special feature of (0.1).

2.1. Assumptions on the Markov process

Let $(\Omega, (\mathscr{F}_t), (Z_t), (\boldsymbol{P}_z)_{z\in E})$ be a conservative Hunt–Markov process valued in a general Polish space E, with a semigroup of transition probability kernels $(P_t)_{t\geqslant 0}$ on E, where

- $\Omega = \boldsymbol{C}(\mathbb{R}^+, E)$ equipped with the compact convergence topology if the process is continuous, or
- $\Omega = \boldsymbol{D}(\mathbb{R}^+, E)$ (the space of *càdlàg mappings* $\omega: \mathbb{R}^+ \to E$) equipped with the Skorokhod topology in the general case;
- $Z_t(\omega) = \omega(t)$ and $\mathscr{F}_t := \sigma(Z_s;\ 0\leqslant s\leqslant t)$ for all $t\geqslant 0$;
- $\boldsymbol{P}_z$ is the law of the Markov process with initial state $z \in E$.

For an initial measure β on E, let $\boldsymbol{P}_\beta(\mathrm{d}\omega) = \int_E \boldsymbol{P}_z(\mathrm{d}\omega)\beta(\mathrm{d}z)$. We write $\boldsymbol{E}^z$ or $\boldsymbol{E}^\beta$ for the expectation under the probability measure $\boldsymbol{P}_z$ or $\boldsymbol{P}_\beta$.

Throughout this section we assume

$$\exists T > 0 \text{ so that } P_T \text{ is strong Feller;} \tag{2.1}$$

and the following *topological transitivity*:

$$\forall \text{ nonempty open subset } O \text{ of } E, \quad \forall z \in E\colon R_1(z, O) := \int_0^{+\infty} \mathrm{e}^{-t} P_t(z, O)\,\mathrm{d}t > 0. \tag{2.2}$$

An immediate consequence of (2.1) and (2.2) is: $P_T R_1(z,\cdot)$, $z \in E$ *are all equivalent and* $P_T(z,\cdot) \ll P_T R_1(z_0,\cdot)$, *for all* $z_0, z \in E$. Indeed assume that $P_T R_1(z_0, A) = R_1 P_T 1_A(z_0) = 0$ for some $z_0 \in E$ and $A \in \mathscr{B}$. Note that $P_T 1_A \geqslant 0$ is continuous by (2.1). If $P_T 1_A$ were not identically zero over E, we would get $R_1 P_T 1_A(z_0) > 0$ by (2.2), a contradiction. Thus $P_T 1_A \equiv 0$ over E, so is $R_1 P_T 1_A$, the desired claim.

The claim above implies that the Markov process $((Z_t), (\boldsymbol{P}_z))$ is *irreducible* w.r.t. $\mu := P_T R_1(z_0,\cdot)$, and the invariant measure, if it exists, is unique (up to a constant factor) and equivalent to μ (see Meyn and Tweedie, 1993; Revuz, 1976).

We say that a measurable function $f: E \to \mathbb{R}$ belongs to the *extended domain* $\boldsymbol{D}_e(\mathscr{L})$ *of the generator* $\mathscr{L}$ *of* (P_t), *if there is a measurable function* $g : E \to \mathbb{R}$ *so that* $\int_0^t |g|(Z_s)\,\mathrm{d}s < +\infty$, $\forall t > 0, \boldsymbol{P}_z$*-a.s. and*

$$f(Z_t) - f(Z_0) - \int_0^t g(Z_s)\,\mathrm{d}s, \quad t\geqslant 0$$

is a càdlàg $\boldsymbol{P}_z$*-local martingale, for all* $z \in E$. *In that case,* $g := \mathscr{L} f$.

2.2. Large deviations

For the language of large deviations, we refer to Deuschel and Stroock (1989), and Dembo and Zeitouni (1998). We begin with several necessary notations and definitions.

Let $M_1(E)$ (resp. $M_b(E)$) be the space of probability measures (resp. signed σ-additive measures of bounded variation) on E equipped with the Borel σ-field $\mathscr{B}$. The usual duality relation between $v \in M_b(E)$ and $f \in b\mathscr{B}$ will be denoted by

$$v(f) := \int f \, \mathrm{d}v.$$

On $M_b(E)$ (or its subspace $M_1(E)$), besides the usual weak convergence topology $\sigma(M_b(E), C_b(E))$, we will consider the *so called τ-topology* $\sigma(M_b(E), b\mathscr{B})$, which is much stronger (see Deuschel and Stroock, 1989; Dembo and Zeitouni, 1998, etc). The σ-field on $M_b(E)$ that we consider in this paper is $\sigma(v \to v(f) | f \in b\mathscr{B}) := \mathscr{M}^\tau$.

On the space $M_1(\Omega)$ of probability measures on Ω, instead of the usual weak convergence topology, we will consider the *projective limit τ-topology* τ_p, generated by $\{Q \to \int F \, \mathrm{d}Q;\ F \in b\mathscr{F}_t,\ \forall t \in \mathbb{R}^+\}$, which is much stronger. The σ-field on $M_1(\Omega)$ generated by $\{Q \to \int F \, \mathrm{d}Q;\ F \in b\mathscr{F}_t,\ \forall t \in \mathbb{R}^+\}$ will be denoted by $\mathscr{M}_p^\tau$. Here $b\mathscr{F}_t$ is the space of all real bounded $\mathscr{F}_t$-measurable functions on Ω.

The level-2 and level-3 empirical measures L_t, R_t given by (0.3) and (0.4), are respectively random elements in $M_1(E)$ and in $M_1(\Omega)$.

The *Donsker and Varadhan level*-3 *entropy functional* $H : M_1(\Omega) \mapsto [0, +\infty]$ is given by

$$H(Q) = \boldsymbol{E}^{\bar{Q}} h_{\mathscr{F}_1^0}(\bar{Q}_{\omega_{(-\infty,0]}}; \boldsymbol{P}_{\omega(0)}) \quad \text{if } Q \in M_1^s(\Omega); +\infty \text{ else} \tag{2.3}$$

where $M_1^s(\Omega)$ is the space of those $Q \in M_1(\Omega)$ which are stationary; $\bar{Q}$ is the unique stationary extension of $Q \in M_1^s(\Omega)$ to $\bar{\Omega} = \boldsymbol{C}(\mathbb{R}, E)$ or $\boldsymbol{D}(\mathbb{R}, E)$, $\bar{Q}_{\omega_{(-\infty,0]}} = \bar{Q}(\cdot | Z_t,\ t \leqslant 0)$ is the regular conditional distribution; and $h_{\mathscr{F}_1^0}(\cdot; \boldsymbol{P}_z)$ is the usual Kullback entropy of $\cdot$ w.r.t. $\boldsymbol{P}_z$ on the σ-field $\mathscr{F}_1^0 = \sigma(Z_t;\ 0 \leqslant t \leqslant 1)$.

The *Donsker and Varadhan level*-2 *entropy functional* $J : M_1(E) \to [0, +\infty]$ is given by

$$J(v) := \inf\{H(Q); Q(Z_0 \in \cdot) = v\}, \quad \forall v \in M_1(E) \tag{2.4}$$

(*Convention:* $\inf \emptyset := +\infty$). See (A.2b) for the Donsker–Varadhan expression of J. Throughout this paper the notation $K \subset\subset E$ means that K is a non-empty compact subset of E.

The proof of the following general result on large deviations will be left to the Appendix:

Theorem 2.1. *Assume* (2.1) *and* (2.2). *Properties* (*a*)–(*d*) *below are equivalent*:

(a) $\boldsymbol{P}_z(L_t \in \cdot)$ *satisfies the LDP on* $M_1(E)$ *w.r.t. the τ-topology with the rate function* J, *uniformly for initial states* z *in the compacts. More precisely*, *the three properties below hold*:

(a.1) J *is inf-compact w.r.t. the τ-topology, i.e.*, $\forall L \geqslant 0$, $[J \leqslant L]$ *is compact in* $(M_1(E), \tau)$;

(a.2) (*the lower bound*) *for any* τ-*open* $\boldsymbol{G} \in \mathscr{M}^\tau$, *and for any* $\boldsymbol{K} \subset\subset E$,

$$\liminf_{t\to+\infty} \frac{1}{t}\log \inf_{z\in \boldsymbol{K}} \boldsymbol{P}_z(L_t \in \boldsymbol{G}) \geqslant -\inf\{J(v);\ v\in \boldsymbol{G}\}; \tag{2.5}$$

(a.3) (*the upper bound*) *for any* τ-*closed* $\boldsymbol{F} \in \mathscr{M}^\tau$, *and* $\boldsymbol{K} \subset\subset E$,

$$\limsup_{t\to+\infty} \frac{1}{t}\log \sup_{z\in \boldsymbol{K}} \boldsymbol{P}_z(L_t \in \boldsymbol{F}) \leqslant -\inf\{J(v);\ v\in \boldsymbol{F}\}; \tag{2.6}$$

(b) $\boldsymbol{P}_z\ (R_t \in \cdot)$ *satisfies the LDP on* $M_1(\Omega)$ *w.r.t. the* τ_p-*topology with the rate function* H, *uniformly for initial states* z *in the compacts.*

(c) $\forall(\lambda > 0,\ \boldsymbol{K}' \subset\subset E)$, *there exists some* $\boldsymbol{K} \subset\subset E$ *such that*

$$\sup_{z\in \boldsymbol{K}'} \boldsymbol{E}^z \exp(\lambda\tau_{\boldsymbol{K}}) < +\infty \quad \textit{and} \quad \sup_{z\in \boldsymbol{K}} \boldsymbol{E}^z \exp(\lambda\tau_{\boldsymbol{K}}(T)) < +\infty \tag{2.7a}$$

where $\tau_{\boldsymbol{K}} = \inf\{t\geqslant 0;\ Z_t \in \boldsymbol{K}\}$ *and* $\tau_{\boldsymbol{K}}(T) = \inf\{t\geqslant T;\ Z_t\in \boldsymbol{K}\}$;

(d) *for any* $\lambda > 0$, *there exists some compact* $\boldsymbol{K} \subset\subset E$ *such that for any* $\boldsymbol{K}' \subset\subset E$,

$$\sup_{z\in \boldsymbol{K}'} \boldsymbol{E}^z \exp(\lambda\tau_{\boldsymbol{K}}(T)) < +\infty. \tag{2.7b}$$

In the case where E *is moreover locally compact, they are equivalent to*

(e) $\boldsymbol{P}_z(L_t \in \cdot)$ *satisfies the LDP on* $M_1(E)$ *w.r.t. the weak convergence topology with the rate function* J, *uniformly for initial states* z *in the compacts of* E.

In each of those cases, (P_t) *has a unique invariant probability measure* α.

Remark. The large deviations of Markov processes are initiated by Donsker and Varadhan in their pioneering works (1975, 1976, 1983), and developed very actively in the last fifteen years by numerous authors. Especially the lower bound of large deviation for any initial measure in the irreducible case without conditions (2.7a) and (2.7b) is established by Ney and Nummelin (1987), de Acosta (1988) in the discrete time case, and by Jain (1990) in the continuous time case (the irreducibility condition is in further removed in Wu (1991, 1992; 2000, Theorem B.1)). For known sufficient conditions to the corresponding upper bounds, see Deuschel and Stroock (1989), de Acosta (1990), Dembo and Zeitouni (1998) and Wu (2000) etc. The reader is referred to Dembo and Zeitouni (1998, Section 6.7) or Deuschel and Stroock (1989) for historical comments and very rich references about large deviations of Markov processes. We are content here only to mention several works which have links with our conditions (2.7a) and (2.7b).

Historically and quite curiously, the studies on the relation between level-2 LDP and recurrence properties began in the opposite direction. Indeed a good candidate for LDP should be the *Doeblin recurrence* (or equivalently uniform exponential recurrence). This was suggested by the important work of Ney and Nummelin (1987) who showed that the level-1 LDP holds for Doeblin recurrent Markov chains (but only for α (the invariant measure)-a.e. initial states!). But four years later Baxter et al. (1991) found for the first time a Doeblin recurrent Markov chain which does not verify the level-2 LDP. In further Bryc and Smolenski (1993) constructed an exponentially recurrent Markov chain for which even the level-1 LDP of $L_t(f)$ for some bounded and measurable f fails. Later Bryc and Dembo (1996) isolated a hyper-exponential α- or ϕ-mixing condition for the LDP for general stationary processes, and found a Doeblin recurrent Markov chain for which the level-2 LDP fails even for the initial measure $\beta=\alpha$ (the invariant measure), showing the sharpness of their conditions.

Condition (2.7b) means that the process is hyper-exponentially recurrent in compact $\boldsymbol{K}$ when $\boldsymbol{K}$ becomes more and more large. It will be called *hyper-exponential recurrence*.

The hyper-exponential recurrence (2.7b) is rather close to the hyper-exponential mixing condition in Bryc and Dembo (1996), intuitively. But mathematically those two conditions are quite different: at first condition (2.7b) is only for Markov processes and theirs is for general stationary processes; in contrast, restricted to the case of Markov processes, their hyper-exponential α-or ϕ-mixing condition is not easy to check in practice and it is stronger than (2.7b) at least in the case where E is countable (by the necessity of (2.7a) and (2.7b) for the LDP).

Theorem 2.1 still holds in the discrete time case by the same proof as that given in the Appendix.

For the upper bound (2.6) above, a key role is played by the Feynman–Kac semigroup

$$P_t^f g(z) := \boldsymbol{E}^z g(Z_t)\cdot \exp\int_0^t f(Z_s)\,\mathrm{d}s \tag{2.8}$$

and the following Cramèr functional:

$$\Lambda_{\boldsymbol{K}}(f) := \limsup_{t\to\infty}\frac{1}{t}\log\sup_{z\in\boldsymbol{K}} P_t^f 1(z). \tag{2.9}$$

Corollary 2.2. *Assume* (2.1) *and* (2.2). *If there is some continuous function* $1\leqslant\Psi\in\boldsymbol{D}_e(\mathscr{L})$ *such that*

$$\phi := -\frac{\mathscr{L}\Psi}{\Psi} \text{ is inf-compact on } E, \tag{C1}$$

then the LDPs in Theorem 2.1 *hold not only uniformly over the compacts*, *but also uniformly over any family of initial measures* $\boldsymbol{A}_\Psi(L) := \{\beta\in M_1(E);\int\Psi\,\mathrm{d}\beta\leqslant L\}$ *where* $L>\inf_E\Psi$ *is arbitrary.*

Moreover for any $v\in M_1(E)$ *with* $J(v)<+\infty$,

$$\int\phi\,\mathrm{d}v\leqslant J(v). \tag{2.10}$$

Proof. We shall verify condition (2.7b). The key remark is that

$$M_t := \Psi(Z_t)\exp\left(\int_0^t\phi(Z_s)\,\mathrm{d}s\right) \text{ is a local martingale,} \tag{2.11}$$

by Ito's formula. Then it is a supermartingale by Fatou's lemma. Consequently, for any $\lambda>0$, taking $A:=\lambda+|\inf_E\phi|+1$ and $\boldsymbol{K}:=[\phi\leqslant A]$ which is compact by (C1), we have for all $t>2T$,

$$\boldsymbol{P}_z(\tau_{\boldsymbol{K}}(T)>t)\leqslant\boldsymbol{P}_z\left(L_t(\phi)\geqslant\frac{T\inf_E\phi+A(t-T)}{t}\right)$$

$$\leqslant \boldsymbol{P}_z\left(\int_0^t \phi(Z_s)\,\mathrm{d}s \geqslant A(t-2T)\right)$$

$$\leqslant \exp(-A(t-2T))\boldsymbol{E}^z \exp\left(\int_0^t \phi(Z_s)\,\mathrm{d}s\right)$$

$$\leqslant \mathrm{e}^{-A(t-2T)}\boldsymbol{E}^z M_t \leqslant \mathrm{e}^{-A(t-2T)}\Psi(z)$$

where (2.7b) follows (since Ψ is continuous on E). Further the previous estimation also implies

$$\sup_{\beta\in \boldsymbol{A}_\Psi(L)} \boldsymbol{E}^\beta \mathrm{e}^{\lambda\tau_K(T)} < +\infty.$$

Proposition A.2 in the Appendix is then applicable and it yields the uniform LDP over $\boldsymbol{A}_\Psi(L)$ for any $L > \inf_E \Psi$.

For the last claim, note at first that $\Lambda_{\boldsymbol{K}}(\phi) \leqslant 0$ ($\Lambda_{\boldsymbol{K}}$ being given by (2.9)), by (2.11). On the other hand, since $v \to \int \phi\,\mathrm{d}v$ is lower semi-continuous w.r.t. the weak convergence topology, by the lower bound of large deviation in Jain (1990) and Laplace principle (Deuschel and Stroock, 1989, Lemma 2.1.7), we have

$$0 \geqslant \Lambda_{\boldsymbol{K}}(\phi) \geqslant \sup\left\{\int \phi\,\mathrm{d}v - J(v);\ v\in[J<+\infty]\right\}$$

where (2.10) follows.

Remark. In the pioneering works Donsker and Varadhan (1975, 1976, 1983), a criterion of approximation type like (C1) (using a sequence (Ψ_n) instead of only one) is given for the LDP of (L_t) w.r.t. the weak convergence topology, under the assumptions well stronger than (2.1) and (2.2). We remark also that under (C1) and the aperiodicity of (P_t), the LDPs in Corollary 2.2 hold in particular for initial measure $\beta = \alpha$ (the invariant measure), because $\alpha(\Psi) < +\infty$ under (C1) by Theorem 2.4 below (since (C1) is stronger than (C2) therein).

For every $f: E \to \boldsymbol{B}$ measurable and bounded where $(\boldsymbol{B}, \|\cdot\|)$ is a separable Banach space, as $v \to \int_E f\,\mathrm{d}v$ is continuous w.r.t. the τ-topology by Deuschel and Stroock (1989, Lemma 3.3.8), then by the contraction principle, $\boldsymbol{P}_z((1/t)\int_0^t f(Z_s)\,\mathrm{d}s \in \cdot)$ satisfies the LDP on $(\boldsymbol{B}, \|\cdot\|)$ with the rate function given by

$$J^f(w) = \inf\left\{J(v) < +\infty | v \in M_1(E) \text{ and } \int f\,\mathrm{d}v = w\right\}, \quad \forall w\in \boldsymbol{B}. \tag{2.12}$$

But what happens for *unbounded* f?

Corollary 2.3. *Assume* (2.1), (2.2) *and* (*C*1). *Given a measurable function* $f: E\to \boldsymbol{B}$ *where* $(\boldsymbol{B}, \|\cdot\|)$ *is a separable Banach space, if there exist* $(A_n)\subset \mathscr{B}$, $(\lambda_n)\subset \mathbb{R}^+$ *and* $(\varepsilon(n)\to 0)$ *such that*

$$\|f\| \leqslant \varepsilon(n)|\phi|\cdot 1_{E\setminus A_n} + \lambda_n 1_{A_n} \tag{2.13}$$

then for any $v\in M_1(E)$ *with* $J(v) < +\infty$, $f\in L^1(E, v)$ *and* $\boldsymbol{P}_\beta((1/t)\int_0^t f(Z_s)\,\mathrm{d}s \in \cdot)$ *satisfies the LDP on* $\boldsymbol{B}$ *with the good rate function* J^f *given by* (2.12), *uniformly w.r.t. initial measures* $\beta \in \boldsymbol{A}_\Psi(L)$ *where* $L > \inf_E \Psi$ *is arbitrary* (*in particular uniformly for initial states* z *in the compacts*).

Proof. The first claim follows from (2.10) and (2.13). We show this LDP in two steps.

Step 1: The mapping $v \to \int f \, \mathrm{d}v$ is continuous on

$$\boldsymbol{F}_a := \left\{ v \in M_1(E); \int \phi \, \mathrm{d}v \leqslant a \right\}$$

for every $a \in \mathbb{R}^+$, where ϕ is given by (C1). Indeed let us consider

$$F^n(v) = \int_{A_n} f \, \mathrm{d}v.$$

By the work of Deuschel and Stroock (1989) F^n is continuous on $(M_1(E), \tau)$ because of the boundedness of $f 1_{A_n}$. Now

$$\sup_{v \in \boldsymbol{F}_a} \left\| F^n(v) - \int f \, \mathrm{d}v \right\| \leqslant \sup_{v \in \boldsymbol{F}_a} \int_{E \setminus A_n} \|f\| \, \mathrm{d}v \leqslant \varepsilon(n) \cdot \sup_{v \in \boldsymbol{F}_a} \int |\phi| \, \mathrm{d}v \leqslant \varepsilon(n) a \to 0$$

as $n \to +\infty$ by our condition (2.13). The desired continuity follows.

Step 2: By Deuschel and Stroock (1989, Exercise 2.1.20) and Step 1, for the LDP in this corollary, we have only to show

$$\lim_{a \to +\infty} \limsup_{t \to +\infty} \frac{1}{t} \log \sup_{\beta \in \boldsymbol{A}_\Psi(L)} \boldsymbol{P}_\beta(L_t \notin \boldsymbol{F}_a) = -\infty$$

for any $L > \inf_E \Psi$. It is very easy. Indeed by Chebychev's inequality and (2.11),

$$\begin{aligned}
\limsup_{t \to +\infty} \frac{1}{t} \log \sup_{\beta \in \boldsymbol{A}_\Psi(L)} \boldsymbol{P}_\beta(L_t \notin \boldsymbol{F}_a) &= \limsup_{t \to +\infty} \frac{1}{t} \log \sup_{\beta \in \boldsymbol{A}_\Psi(L)} \boldsymbol{P}_\beta \left(\int_0^t \phi(Z_s) \, \mathrm{d}s > at \right) \\
&\leqslant \limsup_{t \to +\infty} \frac{1}{t} \log \sup_{\beta \in \boldsymbol{A}_\Psi(L)} \mathrm{e}^{-at} \boldsymbol{E}^\beta \exp \left(\int_0^t \phi(Z_s) \, \mathrm{d}s \right) \\
&\leqslant -a.
\end{aligned}$$

The proof is completed. □

2.3. Exponential convergence

By the strong Feller property (2.1) and the topological transitivity (2.2) (and the resulted irreducibility), every compact subset of E is *petite* in the language of Meyn and Tweedie (1993) and Down et al. (1995). Thus we have

Theorem 2.4 (Down et al., 1995, Theorem 5.2c). *Assume* (2.1), (2.2). *Suppose moreover that our process is aperiodic* (*see Down et al.* (1995) *for definition. That is the case if $P_T(\cdot, K) > 0$ over E for some compact K verifying $P_T R_1(z_0, K) > 0$.*). *If there are some continuous function $1 \leqslant \Psi \in \boldsymbol{D}_e(\mathscr{L})$, some compact subset $\boldsymbol{K} \subset E$ and constants ε, $C > 0$ such that*

$$\phi := -\frac{\mathscr{L}\Psi}{\Psi} \geqslant \varepsilon 1_{\boldsymbol{K}^c} - C 1_{\boldsymbol{K}}, \tag{C2}$$

then there is a unique invariant probability measure α satisfying

$$\int \Psi \, \mathrm{d}\alpha < +\infty, \tag{2.14}$$

and there are some $D>0$ *and* $0<\rho<1$ *such that for all* $t\geqslant 0$,

$$\sup_{|f|\leqslant\Psi}\left|P_tf(z)-\int f\,\mathrm{d}\alpha\right|\leqslant D\Psi(z)\cdot\rho^t\quad\forall z\in E. \tag{2.15}$$

Note. In Down et al. (1995) it is assumed that E is locally compact (certainly for the simplicity of presentation), but their result above does not rely on that assumption. Moreover the extended domain $\boldsymbol{D}_e(\mathscr{L})$ here is slightly larger than that in Down et al. (1995, (12), (13)), but both Theorems 5.1 and 5.2 in Down et al. (1995) hold under our definition of $\boldsymbol{D}_e(\mathscr{L})$, because the key inequality of (Down et al., 1995, (31)) holds under our definition of $\boldsymbol{D}_e(\mathscr{L})$ by Fatou's lemma.

Remark. Let $(B_\Psi,\|\cdot\|_\Psi)$ be the Banach space of all real measurable functions f on E such that

$$\|f\|_\Psi:=\sup_{z\in E}\frac{|f(z)|}{\Psi(z)}<+\infty. \tag{2.16}$$

The exponential convergence (2.15) means that

$$\|(P_t-\alpha)(f)\|_\Psi\leqslant D\rho^t\|f\|_\Psi \tag{2.17}$$

i.e., P_t has a *spectral gap near its largest eigenvalue* 1 *in* B_Ψ.

Corollary 2.5. *Assume* (2.1), (2.2) *and the aperiodicity. Assume that there is some continuous function* $\Psi\geqslant 1$ *satisfying either* (C2) *or* (2.14)+(2.15) (*called* exponential ergodicity *in Down et al.* (1995)). *Let* α *be the unique invariant probability measure. Then*

(a) L_t *converges to* α *with an exponential rate w.r.t. the* τ*-topology. More precisely for any neighborhood* $N(\alpha)$ *of* α *in* $(M_1(E),\tau)$,

$$\sup_{\boldsymbol{K}\subset\subset E}\limsup_{t\to\infty}\frac{1}{t}\log\sup_{z\in\boldsymbol{K}}\boldsymbol{P}_z(L_t\notin N(\alpha))<0. \tag{2.18}$$

(b) *The process is exponentially recurrent in the sense below*: *for any compact* $\boldsymbol{K}$ *in E charged by* α, *there exists some* $\delta>0$ *such that for any compact* $\boldsymbol{K}'$ *in E*,

$$\sup_{z\in\boldsymbol{K}'}\boldsymbol{E}^z\exp(\delta\tau_{\boldsymbol{K}}(T))<+\infty. \tag{2.19}$$

Its proof will be given in the Appendix. The reader is referred to Down et al. (1995) for more informations about exponential ergodicity.

2.4. Moderate deviations

We now turn to moderate deviations. Let $b(t):\mathbb{R}^+\to(0,+\infty)$ be an increasing function verifying

$$\lim_{t\to\infty}b(t)=+\infty,\quad \lim_{t\to\infty}\frac{b(t)}{\sqrt{t}}=0. \tag{2.20}$$

The *moderate deviations of* L_t *from its asymptotic limit* α consist in estimating

$$\boldsymbol{P}_\beta\left(L_t-\alpha\in\frac{b(t)}{\sqrt{t}}A\right)=\boldsymbol{P}_\beta\left(M_t:=\frac{1}{b(t)\sqrt{t}}\int_0^t(\delta_{Z_s}-\alpha)\,\mathrm{d}s\in A\right), \tag{2.21}$$

where A is some $\mathcal{M}^\tau$-measurable subset of $M_b(E)$, a given domain of deviation. When $b(t)=1$, this becomes an estimation of the central limit theorem; and when $b(t)=\sqrt{t}$, it is exactly the large deviation treated in Theorem 2.1. So $b(t)$ satisfying (2.20) is between those two scalings, called *scaling of moderate deviation*.

The spectral gap of (P_t) in B_Ψ in Theorem 2.4 leads to the following moderate deviation principle (in short: MDP) by following Wu (1995) (see the recent work of de Acosta and Chen (1998) in the discrete time case):

Theorem 2.6. *Assume* (2.1), (2.2), *the aperiodicity and* (*C*2). *Then for any initial measure* β *verifying* $\beta(\Psi)<+\infty$, $\boldsymbol{P}_\beta(M_t\in\cdot)$ *satisfies the LDP on* $M_b(E)$ *w.r.t. the* τ-*topology with speed* $b^2(t)$ *and with the rate function given by*

$$I(v):=\sup\left\{\int f\,\mathrm{d}v-\frac{1}{2}\sigma^2(f);\ f\in b\mathscr{B}\right\},\quad \forall v\in M_b(E) \tag{2.22}$$

where

$$\sigma^2(f)=\lim_{t\to\infty}\frac{1}{t}\boldsymbol{E}^\alpha\left(\int_0^t(f-\alpha(f))(Z_s)\,\mathrm{d}s\right)^2 \tag{2.23}$$

exists in $\mathbb{R}$ *for every* $f\in B_\Psi\supset b\mathscr{B}$.

For the moderate deviations of $M_t(f)$ with f unbounded, we have

Theorem 2.7. *Assume* (2.1), (2.2) *and the aperiodicity and* (*C*2). *Given a measurable real function* f *on* E *verifying*

$$\exists M>0 \text{ so that: } |f|\leqslant M\left(\phi-\inf_E\phi+1\right), \tag{2.24}$$

$\boldsymbol{P}_\beta(M_t(f)\in\cdot)$ *satisfies the LDP on* $\mathbb{R}$, *uniformly for* $\beta\in\boldsymbol{A}_\Psi(L):=\{v\in M_1(E);\ v(\Psi)\leqslant L\}$ *for any* $L>\inf_E\Psi$, *with speed* $b^2(t)$ *and with the good rate function* $I^f:\mathbb{R}\to[0,+\infty]$ *given by*

$$I^f(w)=\frac{w^2}{2\sigma^2(f)},\quad \forall w\in\mathbb{R}. \tag{2.25}$$

(*Convention*: $0/0:=0$ *and* $a/0:=+\infty$ *for* $a>0$), *where* $\sigma^2(f)$ *is given by the limit* (2.23).

The proofs of those two results will be given in the Appendix. The MDP in Theorem 2.7 is still valid for $\mathbb{R}^n$-valued f satisfying (2.24), with the rate function given by $I^f(w)=\sup\{w\cdot y-\frac{1}{2}\sigma^2(y\cdot f)\mid y\in\mathbb{R}^n\}$. This can be seen from its proof.

3. Large deviations

We now return to the diffusion (0.1) (under (H1), (H2) and (H3)). About the LDP, we have

Theorem 3.1. *Suppose that there is a function* $G\in C_b^1(\mathbb{R}^d\to\mathbb{R}^d)$ *such that*

$$\nabla V(x)\cdot G(x)\to+\infty\quad \text{as } |x|\to+\infty. \tag{3.1}$$

Assume moreover that there exists some lower bounded $U(x) \in C^2(\mathbb{R}^d)$ *such that*

$$\sup_{x,y\in\mathbb{R}^d} |c^{\mathrm{t}}(x,y)G(x) - \nabla U(x)| < +\infty \tag{3.2}$$

(c^{t} *being the transposition of the matrix* $c(x,y)$). *Then the LDPs in Theorem* 2.1 *hold. Moreover it has a unique invariant probability measure* α *on* $\mathbb{R}^{2d}$ *satisfying*

$$\int_{\mathbb{R}^{2d}} \exp\left[\left(\frac{2c}{\sigma^2} - \delta\right) H(x,y)\right] \mathrm{d}\alpha < +\infty \quad \forall \delta > 0,$$

where $c, \sigma > 0$ *are specified by* $(H2)$ *and* $(H3)$, *respectively.*

Remark 3.2. The last claim above, about the invariant measure α is sharp by the explicit formula (0.2) in the case where $c(x,y) = cI$ and $\Sigma(x,y) = \sigma I$.

Remark that (3.2) is satisfied if $\|c(x,y)\|_{\mathrm{H.S.}}$ is bounded over $\mathbb{R}^{2d}$ (with $U = 0$).

Let us present several particular cases where (3.1) is satisfied:

Case 1: If (0.5) holds, then V satisfies (3.1). Indeed one may take $G(x) = x/|x|$ for $|x| \geqslant 1$ and extend it to the ball $B(o,1) := \{x; |x| < 1\}$ so that $G \in C_b^\infty(\mathbb{R}^d \to \mathbb{R}^d)$.

Case 2: $e_V(x) := \nabla V(x)/|\nabla V(x)| \in C_b^1([|x| > L] \to \mathbb{R}^d)$ *for some* $L > 0$ and

$$\lim_{|x|\to\infty} |\nabla V(x)| = \infty.$$

In that case any $G \in C_b^1(\mathbb{R}^d \to \mathbb{R}^d)$ such that $G(x) = e_V(x)$ for $|x| > L$ satisfies (3.1).

Case 3: $\lim_{|x|\to\infty} |\nabla V(x)| = \infty$ and $e_V(x) := \nabla V(x)/|\nabla V(x)|$ is uniformly continuous outside some ball $B(o,L)$. In this case take $G(x) = (e_V 1_{[|x|>L]}) * h_\varepsilon$ ($*$ is the convolution), where

$$h_\varepsilon(x) = \varepsilon^{-d} h(x/\varepsilon), \quad 0 \leqslant h \in C^\infty, \quad \mathrm{supp}(h) \subset B(o,1), \quad \int h \, \mathrm{d}x = 1.$$

It is easy to see that $G(x)$ verifies (3.1) for $\varepsilon > 0$ small enough.

Case 4: In the unidimensional case ($d = 1$), (0.5), (3.1) and $\lim_{|x|\to\infty} |\nabla V(x)| = \infty$ are all equivalent (under (H1)).

In Section 5, we show the sharpness of (3.1) for the LDP. We now go to the job:

Proof of Theorem 3.1. By Proposition 1.2, our diffusion (0.1) satisfies (2.1), (2.2) and the aperiodicity. By Corollary 2.2, we should construct a Lyapunov function Ψ satisfying (C1). After much tentative, we find a candidate given by

$$\begin{aligned} \Psi = \exp\left(F - \inf_{\mathbb{R}^{2d}} F\right) \quad &\text{where } F(x,y) \\ &= aH(x,y) + (bG(x) + \nabla_x W(x)) \cdot y + bU(x) \end{aligned} \tag{3.3}$$

where G is given in (3.1), U is given in (3.2), and $a, b > 0$ will be determined later as well as some $W \in C_0^2(\mathbb{R}^d)$. F is lower bounded on $\mathbb{R}^{2d}$. As $\Psi \in C^{1,2}(\mathbb{R}^{2d})$, then $\Psi \in \boldsymbol{D}_e(\mathscr{L})$. We have

$$\phi(x,y) = -\frac{\mathscr{L}\Psi}{\Psi} = -\mathscr{L}F - \frac{1}{2}|\Sigma(x,y)\nabla_y F|^2$$

$$= -\frac{a}{2}\mathrm{tr}(\Sigma^2(x,y)) + ac(x,y)y\cdot y - \sum_{i,j=1}^{d} y_i y_j(b\partial_{x_i}G_j(x) + \partial_{ij}W)$$

$$+ y\cdot[bc^{\mathrm{t}}(x,y)G(x) - b\nabla_x U(x) + c^{\mathrm{t}}(x,y)\nabla_x W(x))]$$

$$+ \nabla_x V(x)\cdot(bG(x) + \nabla_x W(x))$$

$$-\frac{1}{2}|\Sigma[ay + bG(x) + \nabla_x W(x)]|^2. \tag{3.4}$$

Now for any $\varepsilon > 0$, by (H3) and Cauchy–Schwartz, the last term above satisfies

$$|\Sigma[ay+bG(x)+\nabla_x W(x)]|^2 \leqslant \sigma^2 \left[a^2(1+\varepsilon)|y|^2 + \left(1+\frac{1}{\varepsilon}\right)|bG(x)+\nabla_x W(x)|^2\right]. \tag{3.5}$$

Fix first any $0 < a < 2c/\sigma^2(1+\varepsilon)$, where $c > 0$ is given by (H2). We choose $b > 0$ so small that

$$b\sup_{x\in\mathbb{R}^d}\|(\partial_{x_i}G_j(x))\|_{\mathrm{H.S.}} < \frac{1}{4}\left(ac - \frac{\sigma^2 a^2(1+\varepsilon)}{2}\right). \tag{3.6}$$

Now choose some function $W(x)\in C_0^2(\mathbb{R}^d)$ with compact support, concave on $B(o,L)$ so that

$$-\left(\frac{b}{2}(\partial_{x_i}G_j(x) + \partial_{x_j}G_i(x)) + \partial_{ij}W\right) \geqslant \left(aA + \frac{a^2\sigma^2(1+\varepsilon)}{2} + 1\right)\cdot I \quad \forall |x|\leqslant L \tag{3.7a}$$

$$-(\partial_{ij}W(x)) \geqslant -\frac{1}{4}\left(ac - \frac{\sigma^2 a^2(1+\varepsilon)}{2}\right)I \quad \forall x\in\mathbb{R}^d \tag{3.7b}$$

where $A > 0$ is some constant so that $c^s(x,y)\geqslant -A\cdot I$ over $\mathbb{R}^{2d}$ (it exists by (H2)). The role of W is to compensate the negative part of $c(x,y)$ for $|x|$ bounded.

With those choices let us see why ϕ is inf-compact on $\mathbb{R}^{2d}$. Indeed by (H2), (3.2) and the assumptions on G, W, there is some constant $M' > 0$ so that

$$y\cdot(bc^{\mathrm{t}}(x,y)G(x) - b\nabla_x U(x) + c^{\mathrm{t}}(x,y)\nabla_x W(x)) \geqslant -M'|y|. \tag{3.8}$$

Hence substituting (3.5) → (3.8) into (3.4), we get

$$\phi(x,y) \geqslant 1_{[|x|>L]}\frac{1}{2}\left(ac - \frac{\sigma^2 a^2(1+\varepsilon)}{2}\right)|y|^2 + 1_{[|x|\leqslant L]}|y|^2 - M'|y|$$

$$+\nabla_x V(x)\cdot(bG(x) + \nabla_x W(x))$$

$$-\frac{\sigma^2}{2}\left(1+\frac{1}{\varepsilon}\right)|bG(x) + \nabla_x W(x)|^2 - \frac{\sigma^2\,\mathrm{d}a}{2}. \tag{3.9}$$

As $W\in C_0^2(\mathbb{R}^d)$ has compact support, $\lim_{|x|+|y|\to\infty}\phi(x,y) = +\infty$ by (3.9) and (3.1). The desired inf-compactness of ϕ follows.

For the last claim, since $\varepsilon > 0$ and $a \in (0, 2c/(\sigma^2(1+\varepsilon)))$ are arbitrary, then for any fixed $\delta > 0$, a can be chosen greater than $2c/\sigma^2 - \delta/2$ in the argument above. Thus the Lyapunov function Ψ in (3.3) satisfies

$$\Psi \geqslant B \exp\left[\left(\frac{2c}{\sigma^2} - \delta\right) H(x, y)\right] \tag{3.10}$$

where $B = B(\delta) > 0$ is some constant. It remains only to apply Theorem 2.4. (2.14). □

Corollary 3.3. *Assume conditions* (3.1) *and* (3.2) *in Theorem* 3.1. *Let* $(B, \|\cdot\|)$ *be a separable Banach space. For any measurable* $f(x, y) : \mathbb{R}^{2d} \to \boldsymbol{B}$ *bounded over the compact subsets of* $\mathbb{R}^d$, *such that*

$$\lim_{(x,y)\to\infty} \frac{\|f(x, y)\|}{|y|^2 + \nabla_x V(x) \cdot G(x)} = 0 \tag{3.11}$$

then for any $v \in M_1(\mathbb{R}^{2d})$ *with* $J(v) < +\infty$, $f \in L^1(E, v)$ *and* $\boldsymbol{P}_z((1/t)\int_0^t f(Z_s)\,\mathrm{d}s \in \cdot)$ *satisfies the LDP on* $\boldsymbol{B}$ *with the good rate function* J^f *given by* (2.12), *uniformly for initial states* z *in the compact subsets of* $\mathbb{R}^{2d}$.

Proof. It follows from Corollary 2.3 because (3.11) implies condition (2.13) by (3.9). □

To complement the previous results, we present

Proposition 3.4. *Without conditions* (3.1) *and* (3.2), *we have always*

(a) *If* $H(Q) < +\infty$, *then under* Q, *the system of coordinates* (x_t, y_t) *satisfies the s.d.e.*

$$\begin{cases} \mathrm{d}x_t = y_t\,\mathrm{d}t \\ \mathrm{d}y_t = \Sigma(x_t, y_t)\,\mathrm{d}W_t - (c(x_t, y_t)y_t + \nabla V(x_t))\,\mathrm{d}t + \beta_t\,\mathrm{d}t \end{cases}$$

where (W_t) *is a* Q*-Brownian motion,* β_t *is some predictable process satisfying*

$$H(Q) = \frac{1}{2}\boldsymbol{E}^Q \int_0^1 |\Sigma^{-1}(x_t, y_t)\beta_t|^2\,\mathrm{d}t < +\infty. \tag{3.12}$$

(b) *If* $J(v) < +\infty$, *then*

$$\int |y|^2\,\mathrm{d}v(x, y) < +\infty \tag{3.13}$$

and

$$\int y\,\mathrm{d}v(x, y) = 0 \quad \text{and} \quad \int \nabla_x f(x) y\,\mathrm{d}v(x, y) = 0 \tag{3.14}$$

for any $f \in C_b^1(\mathbb{R}^d)$.

Proof. Part (a) follows from the Girsanov formula (see Bernard and Wu, 1998 for detail). For part (b), we revisit the proof of Theorem 3.1: choose Ψ as in (3.3), but with $G=0$ and $U=0$. For any $0<a<2c/\sigma^2$, we can find some $W\in C_0^2(\mathbb{R}^d)$ with compact support, so that

$$\phi(x,y):=-\frac{\mathscr{L}\Psi}{\Psi}\geqslant\delta|y|^2-B\quad\forall(x,y)\in\mathbb{R}^{2d}$$

for some constants $B,\delta>0$. On the other hand $\Lambda_{\boldsymbol{K}}(\phi)\leqslant 0$ by (2.11), for any compact $\boldsymbol{K}$ of $\mathbb{R}^{2d}$. This implies by the lower bound of large deviations which holds by Jain (1990, Theorem 4.5) (as in the proof of (2.10)),

$$\text{if } J(v)<+\infty,\quad\text{then } \delta\int|y|^2\,\mathrm{d}v-B\leqslant\int\phi\,\mathrm{d}v\leqslant J(v),$$

where (3.13) follows. To show (3.14), recall that $J(v)$ is the infimum of $H(Q)<+\infty$ among all stationary laws Q on Ω with marginal law v. But for each such Q, we have by part (a) that for every $f\in C_b^1(\mathbb{R}^d)$,

$$0=\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{E}^Q f(x_t)=\boldsymbol{E}^Q\nabla_x f(x_t)y_t=\int\nabla_x f(x)\cdot y\,\mathrm{d}v(x,y).$$

Taking $f_i\in C_b^1$ such that $f_i(x)=x_i$ for $|x_i|\leqslant N$, we obtain also the first equality in (3.14) by letting N go to infinity and by the control (3.13). $\square$

Remark 3.5. Under the conditions of Theorem 3.1, by Corollary 3.3 and (3.14) above we have

$$\limsup_{t\to\infty}\frac{1}{t}\log\sup_{z\in\boldsymbol{K}}\boldsymbol{P}_z\left(\frac{1}{t}\left|\int_0^t y_s\,\mathrm{d}s\right|\geqslant\delta\right)$$
$$\leqslant-\inf\left\{J(v);\left|\int y\,\mathrm{d}v\right|\geqslant\delta\right\}=-\infty\quad\forall\delta>0. \tag{3.15}$$

Since $\int_0^t y_s\,\mathrm{d}s=x_t-x_0$, this seems to be very natural. But in Section 5 we shall show that if (3.1) is not verified, (3.15) may fail!

4. Exponential convergence and moderate deviations

Theorem 4.1. *Suppose that there are a function $G\in C_b^1(\mathbb{R}^d\to\mathbb{R}^d)$ and a lower bounded $U\in C_b^1(\mathbb{R}^d)$ such that $|G(x)|\leqslant 1$ over $\mathbb{R}^d$ and*

$$\liminf_{|x|\to\infty}\nabla V(x)\cdot G(x)>K>0\quad\textit{and}\quad\lim_{|x|\to\infty}\|(\partial_{x_i}G_j(x))\|_{\mathrm{H.S.}}=0 \tag{4.1}$$

$$\lim_{|x|\to\infty}\sup_{y\in\mathbb{R}^d}|c^{\mathrm{t}}(x,y)G(x)-\nabla U(x)|=0. \tag{4.2}$$

Then condition (C2) *in Theorem* 2.4 *is satisfied by some* $1\leqslant\Psi\in C^{1,2}(\mathbb{R}^{2d})$. *In particular, the diffusion* (0.1) *is exponentially ergodic in the sense of* (2.14) + (2.15), *and it satisfies the MDP in Theorem* 2.6.

Moreover for any measurable function $f:\mathbb{R}^{2d}\to\mathbb{R}$ satisfying

$$|f(x,y)|\leqslant M\cdot\left(1+|y|^2+\nabla V(x)\cdot G(x)-\inf_{\mathbb{R}^d}\nabla V\cdot G\right)\quad\forall(x,y)\in\mathbb{R}^{2d} \tag{4.3}$$

for some $M>0$, then $\boldsymbol{P}_z(M_t(f)\in\cdot)$ satisfies the MDP in Theorem 2.7, *uniformly for z in the compacts.*

Remark 4.2. Condition (0.6) given in the Introduction implies (4.1) (by taking $G \in C^\infty(\mathbb{R}^d)$ so that $G(x) = x/|x|$ for $|x| > 1$ and $|G| \leqslant 1$ over $\mathbb{R}^d$). Remark that condition (4.3) for the MDP is much weaker than (3.11) for the LDP.

Since (C1) is stronger than (C2), then all claims of Theorem 4.1 hold if conditions (3.1) and (3.2) are valid, instead of (4.1) and (4.2).

Proof. By Proposition 1.2, the diffusion (0.1) satisfies (2.1), (2.2) and the aperiodicity.

The proof below is close to that of Theorem 3.1. We shall try Ψ given by (3.3) again, except the constants $a, b > 0$ will be chosen differently. For ϕ given by (3.4) to satisfy eventually (C2), using (3.5) for $\varepsilon = 1$, we should ask for

$$bK > \sigma^2 b^2 + \frac{\sigma^2\, \mathrm{d}a}{2} \tag{4.4}$$

($K > 0$ being given by (4.1)) and

$$ac > \sigma^2 a^2. \tag{4.5}$$

We first fix some small $b > 0$ so that $bK > \sigma^2 b^2$. Choose next $a > 0$ so small that both (4.4) and (4.5) hold. By (H2) and the second condition in (4.1), we may find some $L > 0$ so that

$$c^s(x,y) \geqslant cI \quad \text{and} \quad b\|(\partial_{x_i} G_j(x))\|_{\mathrm{H.S.}} \leqslant \frac{1}{4}(ac - \sigma^2 a^2) \quad \forall (|x| > L,\ y \in \mathbb{R}^d). \tag{4.6}$$

Finally take some function $W \in C_0^2(\mathbb{R}^d)$ with compact support such that

$$ac^s(x,y) - \left(\frac{b}{2}(\partial_{x_i} G_j(x) + \partial_{x_j} G_i(x)) + \partial_{ij} W\right) \geqslant (\sigma^2 a^2 + 1)I \quad \forall (|x| \leqslant L, y \in \mathbb{R}^d);$$

$$-(\partial_{ij} W(x)) \geqslant -\frac{1}{4}(ac - \sigma^2 a^2)I \quad \forall x \in \mathbb{R}^d.$$

Substituting all them into (3.4) we get

$$\begin{aligned}\phi(x,y) \geqslant{}& \left(1_{[|x| \leqslant L]} + \frac{1}{2}(ac - \sigma^2 a^2) 1_{[|x| > L]}\right) |y|^2 \\ &- y \cdot (bc^{\mathrm{t}}(x,y) G(x) - b\nabla U(x) + \nabla W(x)) \\ &+ (\nabla V(x) \cdot (bG(x) + \nabla W(x)) - bK) + \left(bK - \sigma^2 b^2 - \frac{\sigma^2\, \mathrm{d}a}{2}\right)\end{aligned} \tag{4.7}$$

where condition (C.2) follows by (4.1) and (4.2). Thus the conclusions in Theorems 2.4 and 2.6 are valid.

For the last claim, if f verifies (4.3), then it satisfies condition (2.24) in Theorem 2.7 by (4.7). $\square$

5. Several examples

5.1. Sharpness of our conditions on the potential

In the unidimensional case, under (H1) for the potential V, condition (3.1) is equivalent to $\lim_{x \to \pm\infty} V'(x) = \pm\infty$; and (4.1) is equivalent to $\liminf_{x \to \pm\infty} \pm V'(x) > 0$.

We show now the sharpness of (3.1) and (4.1) respectively for the LDP and for the exponential convergence:

Proposition 5.1. *Assume that* $d=1$, $c(x,y)\equiv c>0$, $\Sigma(x,y)\equiv\sigma>0$ *and* $V\in C^2(\mathbb{R})$ *is lower bounded.*

(a) *If either*

$$\limsup_{x\to+\infty} V'(x) < +\infty \tag{5.1a}$$

or

$$\liminf_{x\to-\infty} V'(x) > -\infty \tag{5.1b}$$

then all the LDPs in Theorem 2.1 *fail.*

(b) *If either*

$$\limsup_{x\to+\infty} V'(x)\leqslant 0 \tag{5.2a}$$

or

$$\liminf_{x\to-\infty} V'(x)\geqslant 0 \tag{5.2b}$$

then diffusion (0.1) *is not exponentially recurrent in the sense* (2.19) *in Corollary* 2.5. *In particular the exponential convergence* (2.15) *in Theorem* 2.4 *cannot occur.*

Proof. We will examine only the cases (5.1a) and (5.2a), and the cases (5.1b) or (5.2b) can be treated in the same way.

In the case of (5.1a) or of (5.2a), then for *some* $a>0$ or for *all* $a>0$ respectively, we have

$$V'(x)\leqslant a \quad \forall x>L \tag{5.3}$$

where $L=L(a)>0$ is sufficiently large.

Fix $x>L$. On some filtered probability space $(\Omega,(\mathscr{F}_t),\boldsymbol{P})$, let $(W_t)_{t\geqslant 0}$ be a standard Brownian motion w.r.t. $(\mathscr{F}_t)$, and Y_0 a $\mathscr{F}_0$-measurable random variable independent of $(W_t)_t$, whose law $v(\mathrm{d}y)$ is absolutely continuous and has compact support. Let (x_t,y_t) be the unique strong solution of (0.1) with initial condition (x,Y_0). Consider the first hitting time of (x_t) to $(-\infty,L]$:

$$\sigma_L:=\inf\{t\geqslant 0;\ x_t\leqslant L\}.$$

We shall establish that under (5.3), for any $b>0$,

$$\liminf_{n\to\infty}\frac{1}{n}\log \boldsymbol{P}(\sigma_L\geqslant n)\geqslant -\frac{(a+b)^2}{2\sigma^2}\quad \text{for all } x\gg L. \tag{5.4}$$

($x\gg L$ means that $x-L$ is large enough, depending eventually on b). It implies for any $\lambda>a^2/(2\sigma^2)$,

$$\boldsymbol{E}\mathrm{e}^{\lambda\sigma_L}=\infty \quad \text{for all } L>0 \text{ sufficiently large and for all } x\gg L,$$

then (2.7b) is violated. Thus the LDPs in Theorem 2.1 fail. In the case (5.2a), the constant $a>0$ above can be arbitrarily small, then the diffusion is not exponentially recurrent in the sense (2.19). Therefore by Corollary 2.5, the exponential convergence (2.15) fails.

Hence to show both the part (a) and (b), we have only to establish (5.4). We divide its proof into three steps.

Step 1: Let us consider an Ornstein–Uhlenbeck process given by

$$d\tilde{y}_t = \sigma\, dW_t - (c\tilde{y}_t + a)\, dt, \quad \tilde{y}_0 = Y_0. \tag{5.5}$$

whose law on $C(\mathbb{R}^+, \mathbb{R})$ is denoted by $\boldsymbol{P}_\nu^a$. We claim that

$$\tilde{y}_t \leqslant y_t \quad \forall t \leqslant \sigma_L \quad \text{a.s.} \tag{5.6}$$

In fact, $t \to \tilde{y}_t - y_t$ is a process of finite variation and $\forall t \leqslant \sigma_L$

$$(\tilde{y}_t - y_t) = -c\int_0^t (\tilde{y}_s - y_s)\, ds + \int_0^t (V'(x_s) - a)\, ds \leqslant -c\int_0^t (\tilde{y}_s - y_s)\, ds,$$

where (5.6) follows by Gronwall's inequality.

Step 2: for any $\varepsilon, b > 0$, let $\boldsymbol{P}_\nu^{-b}$ be the law of (5.5) with a replaced by $-b$, and consider

$$A_n(x) = \left\{ y_\cdot \in C(\mathbb{R}^+, \mathbb{R});\ \int_0^u y_s\, ds > L - x,\ \forall u \in [0, n] \right\}. \tag{5.7a}$$

$$B_n = \left\{ y_\cdot \in C(\mathbb{R}^+, \mathbb{R});\ \frac{1}{n} \log \left.\frac{d\boldsymbol{P}_\nu^{-b}}{d\boldsymbol{P}_\nu^a}\right|_{\mathscr{F}_n} (y_\cdot) \leqslant \frac{(a+b)^2}{2\sigma^2} + \varepsilon \right\}. \tag{5.7b}$$

By (5.6), we have

$$\begin{aligned} \boldsymbol{P}(\sigma_L \geqslant n) &\geqslant \boldsymbol{P}\left(\int_0^u \tilde{y}_s\, ds > L - x,\ \forall u \in [0, n] \right) = \boldsymbol{P}_\nu^a(A_n(x)) \\ &= \int_{A_n(x)} \exp\left(-\log \left.\frac{d\boldsymbol{P}_\nu^{-b}}{d\boldsymbol{P}_\nu^a}\right|_{\mathscr{F}_n} \right) d\boldsymbol{P}_\nu^{-b} \\ &\geqslant \exp\left[-n\left(\frac{(a+b)^2}{2\sigma^2} + \varepsilon \right) \right] \cdot \boldsymbol{P}_\nu^{-b}(A_n(x) \cap B_n). \end{aligned} \tag{5.7c}$$

Let γ_a (resp. γ_{-b}) be the invariant probability measure of the Ornstein–Uhlenbeck process $\boldsymbol{P}_\cdot^a$ (resp. $\boldsymbol{P}_\cdot^{-b}$), which is Gaussian with mean $-a/c$ (resp. b/c) and with variance σ^2/c (resp. σ^2/c). Put

$$f(y_{[0,1]}) := \log \left.\frac{d\boldsymbol{P}_{y_0}^{-b}}{d\boldsymbol{P}_{y_0}^a}\right|_{\mathscr{F}_1}. \tag{5.8a}$$

We have

$$\log \left.\frac{d\boldsymbol{P}_\nu^{-b}}{d\boldsymbol{P}_\nu^a}\right|_{\mathscr{F}_n} = \sum_{k=0}^{n-1} f(y_{[k,k+1]}). \tag{5.8b}$$

By the Girsanov formula (the calculus is left to the reader),

$$\int f(y_{[0,1]})\, d\boldsymbol{P}_{\gamma_{-b}}^{-b} = \frac{1}{2\sigma^2}(-b-a)^2. \tag{5.9}$$

By Birkhoff's ergodic theorem and (5.8b),

$$\frac{1}{n}\log\left.\frac{\mathrm{d}\boldsymbol{P}_\nu^{-b}}{\mathrm{d}\boldsymbol{P}_\nu^{a}}\right|_{\mathscr{F}_n} \to \frac{(a+b)^2}{2\sigma^2}$$

$\boldsymbol{P}_{\gamma_{-b}}^{-b}$-a.s., then $\boldsymbol{P}_\nu^{-b}$-a.s. (because $\boldsymbol{P}_\nu^{-b} \ll \boldsymbol{P}_{\gamma_{-b}}^{-b}$). Thus we obtain

$$\boldsymbol{P}_\nu^{-b}(B_n) \to 1. \tag{5.10}$$

Step 3: Now for (5.4), by (5.7c) (where $b,\varepsilon > 0$ are arbitrary) and (5.10), it remains to show that for any $b > 0$,

$$\liminf_{n\to\infty} \boldsymbol{P}_\nu^{-b}(A_n(x)) > 0 \quad \forall x \geqslant L. \tag{5.11}$$

Note that

$$A_n(x) \supset A_N(x) \cap \left\{ y.; \int_0^u y_s\,\mathrm{d}s > L-x,\ \forall u \in [N,n] \right\} \supset A_N(x) \cap C_N$$

where $C_N = \{y.; \inf_{u\geqslant N}(1/u)\int_0^u y_s\,\mathrm{d}s \geqslant 0\}$. Note that by the ergodic theorem,

$$\frac{1}{t}\int_0^t y_s\,\mathrm{d}s \to \int y\,\mathrm{d}\gamma_{-b} = \frac{b}{c} > 0, \quad \boldsymbol{P}_{\gamma_{-b}}^{-b}\text{-a.s. then } \boldsymbol{P}_\nu^{-b}\text{-a.s.}$$

Consequently $\lim_{N\to\infty} \boldsymbol{P}_\nu^{-b}(C_N) = 1$. Fix some $N > 0$ so that

$$\boldsymbol{P}_\nu^{-b}(C_N) > \tfrac{3}{4}. \tag{5.12}$$

On the other hand, by the classical maximal ergodic theorem (see, e.g. Revuz (1976) in the discrete time case, which can be extended to the continuous time case by taking the dyadic points approximation, as well known),

$$\boldsymbol{P}_{\gamma_{-b}}^{-b}((A_N(x))^c) \leqslant \boldsymbol{P}_{\gamma_{-b}}^{-b}\left(y. : \inf_{0\leqslant u\leqslant N} \frac{1}{u}\int_0^u y_s\,\mathrm{d}s \leqslant \frac{L-x}{N} \right) \leqslant \frac{N}{x-L}\int |y|\,\mathrm{d}\gamma_{-b}(y).$$

By using the fact that $\boldsymbol{P}_\nu^{-b} \ll \boldsymbol{P}_{\gamma_{-b}}^{-b}$, we get for $x - L > 0$ sufficiently large (depending eventually on $b > 0$),

$$\boldsymbol{P}_\nu^{-b}((A_N(x))^c) < \tfrac{1}{4}. \tag{5.13}$$

Combining (5.12) and (5.13) we obtain

$$\boldsymbol{P}_\nu^{b}(A_n(x)) > \frac{1}{2} \quad \forall n \geqslant 0 \text{ and } \forall x \geqslant L,$$

which yields the desired (5.11). □

Remark 5.2. More strikingly, if (5.1a) or (5.1b) holds, then the LDP (3.15) for $(1/t)\int_0^t y_s\,\mathrm{d}s = (1/t)(x_t - x_0)$ fails. We examine here only the case (5.1a). In this case (5.3) holds for some $a, L > 0$. Assume moreover that Y_0 takes values in $[1,\infty)$. For any $0 < \delta < 1$ fixed, we have by (5.6) that for all $x > L$,

$$\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t}\log \boldsymbol{P}\left(\frac{1}{t}\int_0^t y_s\,\mathrm{d}s > \delta \right) \\
&\quad \geqslant \liminf_{t\to\infty} \frac{1}{t}\log \boldsymbol{P}\left([\sigma_L > t] \cap \left[\frac{1}{t}\int_0^t y_s\,\mathrm{d}s > \delta \right] \right)
\end{aligned}$$

$$\geqslant \liminf_{t\to\infty} \frac{1}{t} \log \boldsymbol{P}(\tilde{y}_s > \delta,\ \forall s \in [0,t])$$

$$=: -A(\delta)$$

where $A(\delta)$ is the lowest spectrum of the operator

$$-\frac{\sigma^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2} + (cy + a)\frac{\mathrm{d}}{\mathrm{d}y}$$

restricted to $L^2((\delta, +\infty), \gamma_a)$ with the Dirichlet boundary condition. Since $0 < A(\delta) < +\infty$, then (3.15) fails.

In other words, the diffusion (0.1) provides a (physical) counter-example for which the exponential recurrence and ergodicity may hold but the LDPs in Theorem 2.1 and the level-1 LDP fail (see Baxter et al. (1991), Bryc and Smolenski (1993) and Bryc and Dembo (1996) for more counter-examples).

5.2. The generalized Duffing oscillator

In this model $d = 1$, $c(x, y) = c > 0$ and $\Sigma(x, y) = \sigma > 0$, and $V(x)$ is a polynomial with leader term $a_{2k}x^{2k}$ where $k \geqslant 1$ and $a_{2k} > 0$. Recall that its unique invariant measure α is given by the explicit expression (0.2).

For $G(x) := \pm 1$ for $\pm x \geqslant 1$, $V'(x)G(x) \cong 2ka_{2k}|x|^{2k-1}$ as $|x|$ goes to infinity. Then the LDPs in Theorem 2.1, the spectral gap in Theorem 2.4 and the MDP in Theorem 2.6 hold all, by Theorems 3.1 and 4.1. But one can do better for the LDP and the MDP of $L_t(f)$ for unbounded f.

Indeed if we choose $G(x) = x$, $U(x) = cx^2/2$, $W = 0$ in definition (3.3) of Ψ with $0 < a < 2c/\sigma^2$ and with $b > 0$ sufficiently small, we will find from (3.4) that $\phi(x, y) := -\mathscr{L}\Psi/\Psi \geqslant \varepsilon(|y|^2 + V'(x)x) - B$ for some constants $\varepsilon, B > 0$. Thus by Corollary 2.3, the LDP for $L_t(f)$ holds for all measurable $\mathbb{R}^n$-valued functions $f(x, y)$ satisfying

$$f \text{ is locally bounded} \quad \text{and} \quad \lim_{|x|+|y|\to\infty} \frac{|f(x,y)|}{H(x,y)} = 0 \tag{5.14}$$

(in particular for $f(x, y) = x$). Moreover by Theorem 2.7, the MDP for $M_t(f)$ holds for all real measurable functions $f(x, y)$ satisfying

$$f \text{ is locally bounded} \quad \text{and} \quad \limsup_{|x|+|y|\to\infty} \frac{|f(x,y)|}{H(x,y)} < +\infty. \tag{5.15}$$

In particular this is valid for $f(x, y) = x$ or $y^2/2$ or $H(x, y)$, main interesting objects in practice.

5.3. The van der Pol model

In this model, $d = 1$, $c(x, y) = cx^2 - c'$ with $c, c' > 0$, $V(x) = \frac{1}{2}\omega_0^2 x^2$ and $\Sigma(x, y) \equiv \sigma > 0$. It seems that its invariant measure is unknown. This model satisfies (H1), (H2) and (H3).

For $G(x) := \pm 1$ and $U(x) := c|x|^3/3 - c'|x|$ for $|x| > 1$ (extended to $|x| \leqslant 1$ smoothly), we see that conditions (3.1) and (3.2) are satisfied. Then the LDP in Theorem 3.1 and the MDP in Theorem 4.1 hold both.

One can get better results on the LDP and MDP of $L_t(f)$ for unbounded f. Indeed given three arbitrary constants $a > 0$, $0 < \varepsilon < 2c\omega_0^2/\sigma^2$ and $N > 0$, we take again the Lyapunov test function Ψ given by the expression (3.3), with the following choices:

- $G(x) = x$, $U(x) = \int G(x)c(x)\,\mathrm{d}x = cx^4/4 - c'x^2/2$; $2\omega_0^2/\sigma^2 - \varepsilon/c < b < 2\omega_0^2/\sigma^2$; and
- $W \in C_0^2(\mathbb{R})$ sufficiently concave on a large ball.

With those choices, $\Psi(x, y) \geqslant \exp(aH(x, y) + (c\omega_0^2/2\sigma^2 - \varepsilon)|x|^4 - A)$ for some constant $A > 0$. Moreover from (3.4), we see easily that for a well chosen W (depending on a, b, c, c', N),

$$\phi(x, y) := -\frac{\mathscr{L}\Psi}{\Psi} \geqslant \delta|x|^2 + ((ac - \varepsilon)|x|^2 + N) \cdot |y|^2 - B$$

for some $0 < \delta < b\omega_0^2 - \sigma^2 b^2/2$ and for some $B > 0$. Then by Theorem 2.4, the unique invariant measure α of this model verifies

$$\int \exp(aH(x, y) + [c\omega_0^2/2\sigma^2 - \varepsilon]|x|^4)\,\mathrm{d}\alpha(x, y) < +\infty \quad \forall a, \varepsilon > 0. \tag{5.16}$$

Moreover, the LDP for $L_t(f)$ holds for $f(x, y) = x$ or for all f satisfying (5.14), and the MDP for $M_t(f)$ is valid for all $f(x, y)$ satisfying (5.15) (in particular for $f(x, y) = x, y^2/2, H(x, y)$), respectively by Corollary 2.3 and Theorem 2.7. Further by the proof of Corollary 2.3 and the estimation of ϕ above, we can show that the LDP of $L_t(f)$ in Corollary 2.3 holds for $f(x, y) = xy$ and for $f(x, y) = y^2/2$ (the kinetic energy).

Acknowledgements

I am grateful to the anonymous referee for his careful and conscientious comments on the first version of this paper, and especially for pointing out a gap in the proof of Step 3 of the implication (c) $\Rightarrow$ (b) in Theorem 2.1 (mentioned therein). My thanks go to my colleague P. Bernard who drew my attention to the model (0.1).

Appendix

Recall that $K \subset\subset E$ means that K is a nonempty compact subset of E.

Proof of Theorem 2.1. Both (b) $\Rightarrow$ (a) and (a) $\Rightarrow$ (e) are trivial by the contraction principle. That (d) $\Rightarrow$ (c) is trivial too, because $\tau_{\boldsymbol{K}} \leqslant \tau_{\boldsymbol{K}}(T)$.

(e) $\Rightarrow$ (d) (in the locally compact case). For any $\lambda > 0$, let $L := \lambda + 1$. The level set $[J \leqslant L]$ is compact in $M_1(E)$ w.r.t. the weak convergence topology. By Prokhorov criterion, for any $\varepsilon > 0$, there exists $\boldsymbol{D} \subset\subset E$ such that $\nu(\boldsymbol{D}^c) < \varepsilon$ for all $\nu \in [J \leqslant L]$. Since E is assumed to be locally compact in this implication, we can find $\boldsymbol{K} \subset\subset E$

so that $\boldsymbol{D}\subset\boldsymbol{K}^{\mathrm{o}}$ (the interior). Noting that $v\to v((\boldsymbol{K}^{\mathrm{o}})^{\mathrm{c}})$ is upper semicontinuous w.r.t. the weak convergence topology, we get by the assumed LDP that for any compact $\boldsymbol{K}'\subset\subset E$,

$$\begin{aligned}
&\limsup_{t\to\infty}\frac{1}{t}\log\sup_{z\in\boldsymbol{K}'}\boldsymbol{P}_z(\tau_{\boldsymbol{K}}(T)>t)\\
&\quad\leqslant\limsup_{t\to\infty}\frac{1}{t}\log\sup_{z\in\boldsymbol{K}'}\boldsymbol{P}_z(L_t(\boldsymbol{K}^{\mathrm{c}})\geqslant 1-\varepsilon)\\
&\quad\leqslant-\inf\{J(v);\ v((\boldsymbol{K}^{\mathrm{o}})^{\mathrm{c}})\geqslant 1-\varepsilon\}\\
&\quad\leqslant-L=\lambda+1,
\end{aligned}$$

where (2.7b) follows.

(a) $\Rightarrow$ (d) (in the general case): We can proceed the proof as above but with $\boldsymbol{K}=\boldsymbol{D}$ simply (by noting that $v\to v(\boldsymbol{K}^{\mathrm{c}})$ is continuous w.r.t. the τ-topology.)

It remains to establish the key implication (c) $\Rightarrow$ (b). Before its proof let us begin with the

Lemma A.1. *Assume*

$$\exists(\mu\in M_1(E),\ S>0)\colon\ \mu P_t\ll\mu\quad\forall t>0\quad\text{and}\quad\mu\ll R_1(z,\cdot),\ P_S(z,\cdot)\ll\mu\quad\forall z\in E. \tag{A.1}$$

(*which is weaker than* (2.1) + (2.2) *as noted at the beginning of Section* 2). *Then for any non-empty compact subset of initial states* $\boldsymbol{K}'\subset E$ *and for all* $v\in M_1(E)$,

$$(\Lambda_{\boldsymbol{K}'})^*(v):=\sup\left\{\int_E f\,\mathrm{d}v-\Lambda_{\boldsymbol{K}'}(f);\ f\in b\mathscr{B}\right\}=J(v)=J^{\mathrm{DV}}(v), \tag{A.2a}$$

where $\Lambda_{\boldsymbol{K}'}$ *is given by* (2.9), *and*

$$J^{\mathrm{DV}}(v):=\sup\left\{\int_E\frac{-\mathscr{L}u}{u}\,\mathrm{d}v;\ 1\leqslant u\in b\mathscr{B}\cap\boldsymbol{D}_e(\mathscr{L}),\ \mathscr{L}u\in b\mathscr{B}\right\}. \tag{A.2b}$$

Proof. For the uniform Cramèr functional Λ_E defined by (2.9) with $\boldsymbol{K}=E$, we know that (see Wu, 2000, Proposition B.13 without the Feller assumption)

$$(\Lambda_E)^*(v)=J(v)=J^{\mathrm{DV}}(v)\quad\forall v\in M_1(E).$$

On the other hand, consider the μ-pointwise upper Cramèr functional

$$\Lambda_\mu^0(f):=\mu-\operatorname*{esssup}_{z\in E}\limsup_{t\to\infty}\frac{1}{t}\log\boldsymbol{E}^z\exp\left(\int_0^t f(Z_s)\,\mathrm{d}s\right)\ \forall f\in b\mathscr{B}$$

where μ is specified by (A.1). We have (see Wu, 2000, Proposition B.10)

$$(\Lambda_\mu^0)^*(v)=J(v)\quad\text{if }v\ll\mu;\ +\infty\text{ otherwise.}$$

Further by Wu (2000, (B.27)) and (A.1), $\Lambda_E(f)\geqslant\Lambda_{\boldsymbol{K}'}(f)\geqslant\Lambda_\mu^0(f)$. Hence to show (A.2a), we have only to prove that $J(v)=+\infty$ if v is not absolutely continuous w.r.t. μ. The last claim is established under (A.1) by Jain (1990). $\square$

(c) $\Rightarrow$ (b): We divide its proof into three steps:

Step 1: (The upper bound (2.6) for L_t). For any fixed $\boldsymbol{K}'\subset\subset E$, let $\Lambda_{\boldsymbol{K}'}(f)$ be the Cramér functional given by (2.9) with $\boldsymbol{K}$ replaced by $\boldsymbol{K}'$. By an extension of the Gärtner–Ellis theorem (see Wu, 1991,1992,2000), if one can prove that

$$\forall (f_n)\subset b\mathscr{B} \text{ decreasing to zero pointwisely on } E\text{: } \Lambda_{\boldsymbol{K}'}(f_n)\to 0, \tag{A.3}$$

then the good upper bound [(a.1)+(a.3)] in Theorem 2.1 holds uniformly for $z\in\boldsymbol{K}'$ with J substituted by the Legendre transformation $(\Lambda_{\boldsymbol{K}'})^*(v)$, defined in (A.2a) above. But by Lemma A.1, $(\Lambda_{\boldsymbol{K}'})^*(v)=J(v)$.

It remains thus to verify (A.3). To this end fix such $(f_n)\subset b\mathscr{B}$. Put $\lambda_0:=2\sup_{z\in E}|f_0(z)|$ and fix some $\lambda\geqslant\lambda_0+1$. Let $\boldsymbol{K}$ be the compact satisfying (2.7a) associated with λ and $\boldsymbol{K}'$. Introduce the successive times of returns to $\boldsymbol{K}$

$$\tau_{\boldsymbol{K}}^0:=\tau_{\boldsymbol{K}},\quad \tau_{\boldsymbol{K}}^{m+1}=\inf\{t\geq\tau_{\boldsymbol{K}}^m+T;\ Z_t\in\boldsymbol{K}\}\quad\forall m\in\mathbb{N}$$

where $T>0$ is specified by our assumption (2.1). We have

$$\int_0^t f_n(Z_s)\,\mathrm{d}s\leqslant\int_0^{\tau_{\boldsymbol{K}}^1} f_n(Z_s)\,\mathrm{d}s+\sum_{k=1}^{[t/T]}\xi_k(n)$$

where $\xi_k(n):=\int_{\tau_{\boldsymbol{K}}^k}^{\tau_{\boldsymbol{K}}^{k+1}} f_n(Z_s)\,\mathrm{d}s$ and $[\,\cdot\,]$ denotes the integer part. Moreover

$$\sup_{z\in\boldsymbol{K}'}\boldsymbol{E}^z\exp\left(\int_0^{\tau_{\boldsymbol{K}}^1} f_n(Z_s)\,\mathrm{d}s\right)\leqslant\sup_{z\in\boldsymbol{K}'}\boldsymbol{E}^z\exp(\lambda\tau_{\boldsymbol{K}}^1)$$
$$\leqslant\sup_{z\in\boldsymbol{K}'}\boldsymbol{E}^z\exp(\lambda\tau_{\boldsymbol{K}})\cdot\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z\exp(\lambda\tau_{\boldsymbol{K}}(T))<+\infty,$$

by (2.7a) and the strong Markov property. Substituting them into (2.9), we get

$$\Lambda_{\boldsymbol{K}'}(f_n)\leqslant\limsup_{m\to\infty}\frac{1}{mT}\log\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z\exp\sum_{k=1}^{m-1}\xi_k(n).$$

To estimate the last expression above, for any $z\in\boldsymbol{K}$, we have by Jensen's inequality and by using the strong Markov property again,

$$\boldsymbol{E}^z\exp\sum_{k=1}^{2m}\xi_k(n)\leqslant\frac{1}{2}\left(\boldsymbol{E}^z\exp\left(\sum_{j=1}^m 2\xi_{2j-1}(n)\right)+\boldsymbol{E}^z\exp\left(\sum_{j=1}^m 2\xi_{2j}(n)\right)\right)$$
$$\leqslant\left(\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z\,\mathrm{e}^{2\xi_1(n)}\right)^m=\left[\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z\exp\left(\int_{\tau_{\boldsymbol{K}}^1}^{\tau_{\boldsymbol{K}}^2}2f_n(Z_s)\,\mathrm{d}s\right)\right]^m.$$

Substituting it into the previous inequality, we get for any $N>T$ fixed,

$$\Lambda_{\boldsymbol{K}'}(f_n)\leqslant\frac{1}{2T}\log\left\{\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z\exp\left(\int_{\tau_{\boldsymbol{K}}^1}^{\tau_{\boldsymbol{K}}^2}2f_n(Z_s)\,\mathrm{d}s\right)\right\}$$
$$\leqslant\frac{1}{2T}\log\left\{\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z\exp\left(\int_{\tau_{\boldsymbol{K}}^1}^{\tau_{\boldsymbol{K}}^1+N}2f_n(Z_s)\,\mathrm{d}s\right)\right.$$

$$\left. + \sup_{z\in \boldsymbol{K}} \boldsymbol{E}^z\, 1_{[\tau_{\boldsymbol{K}}^2 > \tau_{\boldsymbol{K}}^1 + N]} \exp(\lambda_0(\tau_{\boldsymbol{K}}^2 - \tau_{\boldsymbol{K}}^1)) \right\}. \tag{A.4}$$

Having this inequality, we can now prove (A.3) rather easily. Indeed, by condition (2.7a) and our choice $\lambda \geqslant \lambda_0 + 1$, the family $\{\exp(\lambda_0 \tau_{\boldsymbol{K}}(T)), \boldsymbol{P}_z\}_{z\in\boldsymbol{K}}$ is uniformly integrable. Thus

$$\varepsilon(N) := \sup_{z\in\boldsymbol{K}} \boldsymbol{E}^z 1_{[\tau_{\boldsymbol{K}}^2 > \tau_{\boldsymbol{K}}^1 + N]} \exp(\lambda_0(\tau_{\boldsymbol{K}}^2 - \tau_{\boldsymbol{K}}^1)) = \sup_{z\in\boldsymbol{K}} \boldsymbol{E}^z 1_{[\tau_{\boldsymbol{K}}(T) > N]} \exp(\lambda_0 \tau_{\boldsymbol{K}}(T)) \to 0$$

as $N \to \infty$. On the other hand for $z \in \boldsymbol{K}$, $\tau_{\boldsymbol{K}}^1 = \tau_{\boldsymbol{K}}(T)$, then for any $N > T$ fixed,

$$\boldsymbol{E}^z \exp \int_{\tau_{\boldsymbol{K}}^1}^{\tau_{\boldsymbol{K}}^1 + N} 2 f_n(Z_s)\, \mathrm{d}s = (P_T F_n)(z) \quad \text{where } F_n(\cdot) := \boldsymbol{E}^{\cdot} \exp \int_{\tau_{\boldsymbol{K}}}^{\tau_{\boldsymbol{K}} + N} 2 f_n(Z_s)\, \mathrm{d}s.$$

$\{F_n\}$ is bounded by $\mathrm{e}^{\lambda N}$, decreasing to 1 pointwise on E by dominated convergence. Thus

$P_T F_n(z)$ decreases to 1 pointwise over $\boldsymbol{K}$.

By Dini's monotone convergence theorem and the strong Feller property (2.1),

$$\sup_{z\in\boldsymbol{K}} |P_T F_n(z) - 1| \to 0, \quad \text{as } n \to +\infty.$$

Substituting this estimation into (A.4), we obtain

$$\limsup_{n\to+\infty} \Lambda_{\boldsymbol{K}'}(f_n) \leqslant \frac{1}{2T} \log(1 + \varepsilon(N)).$$

Since $\varepsilon(N) \to 0$ as $N \to +\infty$, (A.3) follows.

Step 2 (Lower bound for L_t). Jain (1990, Theorem 4.5) proved the pointwise lower bound of large deviation (for the weak convergence topology; but as indicated in his Remarks 1.3, that holds for the τ-topology), with rate function J for any initial state $z \in E$, under the assumption (A.1). Now the desired locally uniform lower bound (2.5) follows from his general result, the locally uniform good upper bound shown in Step 1, and from the

Claim. *Assume* (A.1). *Let* $\boldsymbol{A}$ *be a nonempty family of initial measures. If* $\boldsymbol{P}_\beta(L_t \in \cdot)$ *satisfies the uniform upper bound of large deviation over* $\beta \in \boldsymbol{A}$, *with good rate function J for the* τ*-topology*, *then it satisfies the corresponding uniform lower bound of large deviation for initial measures* $\beta \in \boldsymbol{A}$.

It can be proven by following Wu (2000) (the last part of the proof of Theorem 5.1), and hence omitted.

Step 3 (The LDP for (R_t)). For every $s > 0$, consider the Markov process $(Y_t^{(s)}(\omega) := \omega_{[t,t+s]} \in \boldsymbol{D}([0,s]; E) =: \Omega_{[0,s]})$. Its transition semigroup $(P_t^{(s)})_{t\geqslant 0}$ satisfies still (2.1) but with $T + s$ instead of T. However it does not necessarily satisfies (2.2) (thinking of the case where the support of $\boldsymbol{P}_\mu|_{\Omega_{[0,s]}}$ with $\mu = P_T R_1(z_0,\cdot)$ is not the whole $\Omega_{[0,s]}$). Nevertheless $(P_t^{(s)})$ is irreducible w.r.t. $\mu^{(s)} := \boldsymbol{P}_\mu|_{\Omega_{[0,s]}}$ with $\mu = P_T R_1(z_0,\cdot)$, and $P_{T+s}^{(s)}(Y_0^{(s)},\cdot) \ll \mu^{(s)}$ for all $Y_0^{(s)} \in \Omega_{[0,s]}$ (easy). In other words $(P_t^{(s)})$ satisfies (A.1).

Notice that if (2.7a) holds for some $T>0$, then it holds for all $S>0$ (in particular for $T+s$). In fact for any $S>0$, then $\tau_{\boldsymbol{K}}(S)\leqslant\tau_{\boldsymbol{K}}^k$ where $k\geqslant S/T$ and $\tau_{\boldsymbol{K}}^k$ is defined in Step 1. Thus we have by the strong Markov property

$$\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z e^{\lambda\tau_{\boldsymbol{K}}(S)}\leqslant\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z e^{\lambda\tau_{\boldsymbol{K}}^k}\leqslant\left(\sup_{z\in\boldsymbol{K}}\boldsymbol{E}^z e^{\lambda\tau_{\boldsymbol{K}}(T)}\right)^k<+\infty$$

the desired claim. As Lemma A.1 remains valid for $(Y_t^{(s)})$ (since it satisfies condition (A.1)), by the same proof as that in Steps 1 and 2 except that the successive times $\tau_{\boldsymbol{K}}^k$, $k\geqslant 1$ of returns to $\boldsymbol{K}$ should be defined now with T substituted by $T+s$, the empirical measure

$$L_t^{(s)}:=\frac{1}{t}\int_0^t\delta_{Y_a^{(s)}}\,da$$

of $(Y_t^{(s)})$ satisfies the LDP w.r.t. the τ-topology on the space $M_1(\Omega_{[0,s]})$ of probability measures on $(\Omega_{[0,s]},\mathscr{F}_s)$, with the rate function $J^{(s)}$ given by (2.4) but associated with $(Y_t^{(s)})$, uniformly for initial states z in the compacts.

Let us identify the level-2 rate function $J^{(s)}$ for $(Y_t^{(s)})$. To that end let $\Omega^{(s)}:=\boldsymbol{D}(\mathbb{R}^+,\Omega_{[0,s]})$ be the path space of the process $(Y_\cdot^{(s)})$ and $H^{(s)}:M_1(\Omega^{(s)})\mapsto[0,+\infty]$ be the level-3 Donsker and Varadhan entropy functional defined by (2.3) but associated with $(Y_t^{(s)})$.

Let $Q^{(s)}\in[H^{(s)}<+\infty]$. It is easy to check that $Q^{(s)}([Y^{(s)}(\Omega)]^c)=0$. Let Q be the image measure of $Q^{(s)}$ by the inverse application $(Y^{(s)})^{-1}:Y^{(s)}(\Omega)\to\Omega$. Then Q is stationary as same as $Q^{(s)}$, and $Q^{(s)}$ is the image measure of Q under the application $Y^{(s)}$. By definition (2.3) (recalling that $\bar{Q}$ is the unique stationary measure on the enlarged space $\bar{\Omega}:=\boldsymbol{D}(\mathbb{R},E)$ extending Q),

$$\begin{aligned}H^{(s)}(Q^{(s)})&=\boldsymbol{E}^{\bar{Q}}h(\bar{Q}(Y_{[0,1]}^{(s)}\in\cdot\,|Y_{(-\infty,0]}^{(s)});\ \boldsymbol{P}_{\omega_{[0,s]}}(Y_{[0,1]}^{(s)}\in\cdot))\\&=\boldsymbol{E}^{\bar{Q}}h(\bar{Q}(\omega_{[s,s+1]}\in\cdot\,|\omega_{(-\infty,s]});\ \boldsymbol{P}_{\omega(s)}(\omega_{[s,s+1]}\in\cdot))\\&=H(Q)\end{aligned}$$

where $\boldsymbol{P}_{\omega_{[0,s]}}(Y_{[0,1]}^{(s)}\in\cdot)$ is the law of $Y_{[0,1]}^{(s)}:=(Y_t^{(s)})_{0\leqslant t\leqslant 1}$ starting from $Y_0^{(s)}(\omega)=\omega_{[0,s]}\in\Omega_{[0,s]}$ associated with $(P_t^{(s)})$ (well defined for every $\omega_{[0,s]}\in\Omega_{[0,s]}$), and $\bar{Q}(Y_{[0,1]}^{(s)}\in\cdot\,|Y_{(-\infty,0]}^{(s)})$ is the regular conditional distribution of $Y_{[0,1]}^{(s)}$ knowing $Y_{(-\infty,0]}^{(s)}$ under $\bar{Q}$.

Then by definition (2.4) of $J^{(s)}$ associated with $(Y_t^{(s)})$, for any $Q_{[0,s]}\in M_1(\Omega_{[0,s]})$,

$$\begin{aligned}J^{(s)}(Q_{[0,s]}):=&\inf\{H^{(s)}(Q^{(s)})\,|\,Q^{(s)}\in M_1(\Omega^{(s)}),\ Q^{(s)}(Y_0^{(s)}\in\cdot)\\&=Q_{[0,s]},\ H^{(s)}(Q^{(s)})<+\infty\}\\=&\inf\{H(Q)\,|\,Q\in M_1^s(\Omega),\ Q(Y_0^{(s)}\in\cdot)=Q_{[0,s]}\},\end{aligned}$$

where $M_1^s(\Omega)$ is the space of all stationary measures on Ω.

Having all those ingredients let us consider the level-3 LDP of (R_t). Since for all $s>0$, $(L_t^{(s)})$ satisfies the LDP with rate function $J^{(s)}$ uniformly for z in the compacts, and since $(M_1(\Omega),\tau_p)$ is the projective limit space of $\{(M_1(\Omega_{[0,s]}),\tau),s\to\infty\}$, by the projective limit theorem of LDP due to Dawson and Gartner (1987) (the reader can extend it easily for the uniform type LDP), $\boldsymbol{P}_z(R_t \in \cdot)$ satisfies the LDP on $(M_1(\Omega), \tau_p)$ uniformly for z in compacts, with the rate function given by

$$\tilde{H}(Q) := \sup_{s>0} J^{(s)}(Q_{[0,s]}),$$

where $Q_{[0,s]}$ is the marginal law of Q on $\Omega_{[0,s]}$. It remains to check that $\tilde{H}(Q) = H(Q)$ over $M_1(\Omega)$.

It is obvious that $\tilde{H}(Q) \leqslant H(Q)$ by the identification of $J^{(s)}$ above. For the inverse inequality, we can assume that $\tilde{H}(Q) < +\infty$ (trivial otherwise). For any $s > 0$, $Q_{[0,s]}$ must be shift-invariant (that means $\int F(\omega_{S+t})\,\mathrm{d}Q_{[0,s]}(\omega) = \int F(\omega_S)\,\mathrm{d}Q_{[0,s]}(\omega)$ for any finite subset S of $[0,s]$ and for all t so that $S + t := \{u + t;\ u \in S\}$ is contained in $[0,s]$), then Q is stationary. In further, for any $s \geqslant 1$, by the convexity of $x \log x$ we have

$$J^{(s)}(Q_{[0,s]}) = \inf\{H(Q') \,|\, Q' \in M_1^s(\Omega),\ Q'(Y_0^{(s)} \in \cdot) = Q_{[0,s]}\}$$
$$\geqslant h_{\mathscr{F}_1^{-s+1}}(\bar{Q}';\ \bar{Q}' \otimes_0 \boldsymbol{P}_\cdot)$$

where $\bar{Q}' \otimes_0 \boldsymbol{P}_\cdot(d\omega_{(-\infty,0]}, d\omega_{[0,+\infty)}) := \bar{Q}'(d\omega_{(-\infty,0]})\boldsymbol{P}_{\omega(0)}(d\omega_{[0,+\infty)})$, and $\mathscr{F}_t^s = \sigma(\omega(a);\ s \leqslant a \leqslant t)$ (on $\bar{\Omega}$). But $h_{\mathscr{F}_1^{-s+1}}(\bar{Q}';\ \bar{Q}' \otimes_0 \boldsymbol{P}_\cdot) = h_{\mathscr{F}_1^{-s+1}}(\bar{Q};\ \bar{Q} \otimes_0 \boldsymbol{P}_\cdot)$ for any $Q' \in M_1^s(\Omega)$ such that $Q'(Y_0^{(s)} \in \cdot) = Q_{[0,s]}$. As s goes to infinity, the last term above tends to (by the submartingale convergence)

$$h_{\mathscr{F}_1^{-\infty}}(\bar{Q};\ \bar{Q} \otimes_0 \boldsymbol{P}_\cdot)$$

which is another expression of $H(Q)$. Hence $\tilde{H}(Q) \geqslant H(Q)$. (Note: the identification of $J^{(s)}$ and the equality $\tilde{H} = H$ here present no real novelty w.r.t. the work in Deuschel and Stroock (1989, Theorem 4.4.38). However their result cannot be applied directly here because they used (loosely) their dominating measure assumption $(\tilde{U})$. This is indicated by the referee.) The proof of (c) $\Rightarrow$ (b) is finished.

For the last claim, we have only to show the existence of the invariant measure α. Indeed by the upper bound of LD, we have

$$0 = \limsup_{t\to\infty} \frac{1}{t} \log \boldsymbol{P}_z(L_t(E) = 1) \leqslant -\inf_{v \in M_1(E)} J(v).$$

The inf-compacity of J implies the existence of $\alpha \in M_1(E)$ satisfying $J(\alpha) = 0$. But the last property is equivalent to the invariance of α w.r.t. (P_t), as well known (Deuschel and Stroock, 1989). $\square$

The proof of Theorem 2.1 yields the following

Proposition A.2. *Assume* (2.1) *and* (2.2) (*or* (*A*.1) *instead of* (2.2)). *Let* $\boldsymbol{A}$ *be a nonempty set of initial measures. If for any* $\lambda > 0$, *there exists some* $\boldsymbol{K} \subset\subset E$ *so that*

$$\sup_{\beta \in \boldsymbol{A}} \boldsymbol{E}^\beta e^{\lambda \tau_{\boldsymbol{K}}} < +\infty, \quad \sup_{z \in \boldsymbol{K}} \boldsymbol{E}^z e^{\lambda \tau_{\boldsymbol{K}}(T)} < +\infty,$$

then $\boldsymbol{P}_\beta(L_t \in \cdot)$ (*resp.* $\boldsymbol{P}_\beta(R_t \in \cdot)$) *satisfies the LDP on* $(M_1(E), \tau)$ (*resp.* $(M_1(\Omega), \tau_p)$), *with rate function* J (*resp.* H), uniformly *for* $\beta \in \boldsymbol{A}$.

Proof of Theorem 2.7. (By following Wu, 1995, Theorem 2.1) Consider the Feynman-Kac semigroup

$$P_t^{\lambda f} g(z) := \boldsymbol{E}^z g(Z_t) \exp\left(\int_0^t \lambda f(Z_s)\, \mathrm{d}s \right) \quad \forall g \in B_\Psi.$$

We consider the complexification of B_Ψ (introduced in (2.16)), denoted still by B_Ψ. When $\lambda \in \boldsymbol{C}, |\lambda| \leqslant \delta$ with $\delta > 0$ sufficiently small, $|\lambda f| \leqslant \phi - \inf_E \phi + 1$, by our condition (2.24).

Letting $A := -\inf_E \phi + 1$ and $|\lambda| \leqslant \delta$, we get by (2.11),

$$\|P_t^{\lambda f} g\|_\Psi \leqslant \|g\|_\Psi \sup_{z \in E} \frac{1}{\Psi(z)} \mathrm{e}^{At} (P_t^\phi \Psi)(z) \leqslant \|g\|_\Psi \mathrm{e}^{At}. \tag{A.5}$$

From this estimation we deduce easily that $\lambda \to P_t^{\lambda f} g \in B_\Psi$ is holomorphic on the disk $D_\delta := \{\lambda \in \boldsymbol{C}; |\lambda| < \delta\}$ (in the usual sense, see Kato, 1984, Chapter VII, p. 365).

Now by Theorem 2.4, 1 is an isolated, simple and the only eigenvalue with modulus 1 of the operator $P_1 = P_1^{\lambda f}$ with $\lambda = 0$ acting on B_Ψ. Let $J(0)f := \langle f, 1 \rangle_\alpha$ be the spectral projection associated with the eigenvalue 1 of P_1. Then the spectral radius of $P_1(I - J(0))$ is not greater than ρ (< 1) appeared in (2.15).

Applying Kato's holomorphic perturbation theorem (Kato, 1984, Chapter VII, Theorems 1.7 and 1.8, p. 368–370), for some $r \in (\rho, (1+\rho)/2)$ and for any $L > \inf_E \Psi$, there is some small disk D_{δ_0} in $\boldsymbol{C}$ with $\delta_0 \in (0, \delta)$ such that for all $\lambda \in D_{\delta_0}$:

- the point $G(\lambda)$ with largest modulus in the spectrum of $P_1^{\lambda f}$ acting on B_Ψ is an isolated eigenvalue, $\lambda \to G(\lambda)$ is holomorphic on D_{δ_0} and $|G(\lambda)| > (1+\rho)/2$;
- the spectral projection $J(\lambda)$ of $P_1^{\lambda f}$ associated with $G(\lambda)$ is also holomorphic and one dimensional, and $\|J(\lambda)1 - J(0)1\|_\Psi = \|J(\lambda)1 - 1\|_\Psi < 1/2L$;
- the spectral radius of $P_1^{\lambda f}(I - J(\lambda))$ is strictly less than r;
- and

$$M := \sup_{z \in S(1/r), \lambda \in D_{\delta_0}} \|(I - zP_1^{\lambda f}[I - J(\lambda)])^{-1}\|_{B_\Psi \to B_\Psi} < +\infty,$$

where $S(a) = \{z \in \boldsymbol{C};\ |z| = a\}$.

Now from the following Cauchy integral formula

$$\begin{aligned}(P_1^{\lambda f}[I - J(\lambda)])^n &= \frac{1}{n!} \frac{\partial^n}{\partial z^n} (I - zP_1^{\lambda f}[I - J(\lambda)])^{-1}|_{z=0} \\ &= \frac{1}{2\pi \mathrm{i}} \int_{S(1/r)} \frac{(I - zP_1^{\lambda f}[I - J(\lambda)])^{-1}}{z^{n+1}} \mathrm{d}z,\end{aligned}$$

we deduce that for all $\lambda \in D_{\delta_0}$ and for all $n \geqslant 1$,

$$\|P_n^{\lambda f} - G(\lambda)^n J(\lambda)\|_{B_\Psi \to B_\Psi} = \|(P_1^{\lambda f}[I - J(\lambda)])^n\|_{B_\Psi \to B_\Psi} \leqslant M r^n. \tag{A.6a}$$

We now extend this crucial estimation for all $t \in \mathbb{R}^+$. As $P_t^{\lambda f}$ commutes with $P_1^{\lambda f}$, it commutes with its spectral projection $J(\lambda)$ too. As $J(\lambda)$ is one dimensional, that implies $P_t^{\lambda f} J(\lambda) = h(t) J(\lambda)$. By the semigroup property and the continuity of $t \to \int P_t^{\lambda f} J(\lambda) 1\, \mathrm{d}\alpha$ (*warning*: the semigoup $(P_t^{\lambda f})$ is however not strongly continuous in B_Ψ in general), there is a constant $a \in \boldsymbol{C}$ such that $h(t) = \mathrm{e}^{at}$ for all $t \in \mathbb{R}^+$. Thus letting $t = 1$, we obtain $a = \log G(\lambda)$. Consequently for all $t \in [n, n+1]$, we have by (A.5) and (A.6a),

$$\begin{aligned} \|P_t^{\lambda f} - \exp(t \log G(\lambda)) J(\lambda)\|_{B_\Psi \to B_\Psi} &= \|P_t^{\lambda f}(I - J(\lambda))\|_{B_\Psi \to B_\Psi} \\ &\leqslant \sup_{0 \leqslant s \leqslant 1} \|P_s^{\lambda f}\|_{B_\Psi \to B_\Psi} \cdot \|P_n^{\lambda f}(I - J(\lambda))\|_{B_\Psi \to B_\Psi} \\ &\leqslant e^A M r^n \leqslant \frac{e^A M}{r} r^t. \end{aligned} \tag{A.6b}$$

Let us consider the family of initial measures

$$A_\Psi(L) = \{\beta \in M_1(E);\ \beta(\Psi) \leqslant L\}.$$

By (A.6b) and the four points above, for all t large enough, $\log \int P_t^{\lambda f} 1 \, d\beta$ are holomorphic on D_{δ_0} for all $\beta \in A_\Psi(L)$ (as same as $\log G(\lambda)$). Moreover we have

$$\lim_{t \to +\infty} \sup_{\lambda \in D_{\delta_0}} \sup_{\beta \in A_\Psi(L)} \left| \frac{1}{t} \log \int P_t^{\lambda f} 1 \, d\beta - \log G(\lambda) \right| = 0. \tag{A.7a}$$

By Cauchy's theorem for holomorphic functions, this implies in further that

$$\sup_{|\lambda| \leqslant \varepsilon} \sup_{\beta \in A_\Psi(L)} \left| \frac{d^k}{d\lambda^k} \frac{1}{t} \log \int P_t^{\lambda f} 1 \, d\beta - \frac{d^k}{d\lambda^k} \log G(\lambda) \right| \to 0 \quad \forall k \in \mathbb{N} \tag{A.7b}$$

as $t \to \infty$, where $\varepsilon \in (0, \delta_0)$ is arbitrary.

Since $\alpha \in A_\Psi(L)$ for some $L > 0$ (by Theorem 2.4), we have by (A.7b),

$$\frac{d^2}{d\lambda^2} \frac{1}{t} \log \int P_t^{\lambda f} 1 \, d\alpha|_{\lambda=0} = \boldsymbol{E}^\alpha \frac{1}{t} \left(\int_0^t (f - \alpha(f))(Z_s) \, ds \right)^2 \to \frac{d^2}{d\lambda^2} \log G(\lambda)|_{\lambda=0},$$

then the limit (2.23) exists. Applying Wu (1995, Theorem 1.2) whose conditions are satisfied by (A.7b), we get the desired MDP.

Proof of Theorem 2.6. For any $f \in b\mathscr{B} \subset B_\Psi$ and for any initial measure satisfying $\beta(\Psi) < +\infty$, by (A.7a) and (A.7b), the exponential convergence (2.15) and by Wu (1995, Theorem 1.2), we have

$$\lim_{t \to \infty} \left| \frac{1}{b^2(t)} \log \boldsymbol{E}^\beta \exp(b^2(t) M_t(f)) - \frac{1}{2} \sigma^2(f) \right| = 0. \tag{A.8}$$

On the other hand by Theorem 2.4,

$$\int_0^\infty |P_t f - \alpha(f)|(z) \, dt \leqslant D \|f\|_{B_\Psi} \Psi(z) \int_0^\infty \rho^t \, dt \quad \text{where} \quad \int_0^\infty \rho^t \, dt < +\infty,$$

then $R_0 f := \int_0^{+\infty} (P_t f - \alpha(f)) \, dt$ is absolutely convergent in $(B_\Psi, \|\cdot\|_\Psi)$ and then in $L^1(\alpha)$ (as $\Psi \in L^1(\alpha)$). Consequently

$$\frac{1}{t} \boldsymbol{E}^\alpha \left(\int_0^t (f(Z_s) - \alpha(f)) \, ds \right)^2 = \frac{2}{t} \int_0^t ds \int_0^s \langle f, P_u f - \alpha(f) \rangle_\alpha \, du \to 2 \int_E f R_0 f \, d\alpha.$$

as t tends to infinity. We obtain so

$$\sigma^2(f) = 2 \int_E f R_0 f \, d\alpha. \tag{A.9}$$

This functional is Gateaux differentiable on $b\mathscr{B}$, and it satisfies

$$\forall (f_n) \subset b\mathscr{B} \text{ uniformly bounded and } f_n(z) \to 0, \quad \forall z \in E, \text{ then } \sigma^2(f_n) \to 0.$$

Those two properties imply the desired LDP of $\boldsymbol{P}_\beta(M_t \in \cdot)$ on $M_b(E)$ w.r.t. the τ-topology, as in the proof of Wu (1995, Theorem 2.3). □

Proof of Corollary 2.5. (a) Since any neighborhood $N(\alpha)$ of α in $(M_1(E), \tau)$ contains a set of form $\{v; |\int f \,\mathrm{d}v - \int f \,\mathrm{d}\alpha| < \varepsilon\}$, where $f = (f_1, \ldots, f_n) \in b\mathscr{B}^n$, $n \geqslant 1$ and $\varepsilon > 0$. Then for (2.18), it is enough to show that

$$\sup_{\boldsymbol{K} \subset\subset E} \limsup_{t \to \infty} \frac{1}{t} \log \sup_{z \in \boldsymbol{K}} \boldsymbol{P}_z(|L_t(f) - \alpha(f)| \geqslant \varepsilon) < 0. \tag{A.10}$$

By Ellis (1985, Theorem II.6.3) (more precisely by its proof), for (A.10) we have only to prove the Gateaux differentiability of $\boldsymbol{u} \to \sup_{\boldsymbol{K} \subset\subset E} \Lambda_{\boldsymbol{K}}(\boldsymbol{u} \cdot f)$ at $\boldsymbol{u} = 0 \in \mathbb{R}^n$. But by (A.7a), for any direction $\upsilon \in \mathbb{R}^n$ and for any nonempty compact $\boldsymbol{K}$, $\Lambda_{\boldsymbol{K}}(\lambda\upsilon \cdot f) = \log G(\lambda)$, which is independent of $\boldsymbol{K}$ and analytic on a neighborhood of $\lambda = 0$, by (A.7b).

(b) For any compact $\boldsymbol{K}$ in E with $\alpha(\boldsymbol{K}) > 0$, by (A.10), we have for any $0 < \varepsilon < 1 - \alpha(\boldsymbol{K})$,

$$\begin{aligned} &\sup_{\boldsymbol{K}' \subset\subset E} \limsup_{t \to \infty} \frac{1}{t} \log \sup_{z \in \boldsymbol{K}'} \boldsymbol{P}_z(\tau_{\boldsymbol{K}}(T) > t) \\ &\quad \leqslant \sup_{\boldsymbol{K}' \subset\subset E} \limsup_{t \to \infty} \frac{1}{t} \log \sup_{z \in \boldsymbol{K}'} \boldsymbol{P}_z(L_t(\boldsymbol{K}) \geqslant 1 - \varepsilon) < 0 \end{aligned}$$

where (2.19) follows.

References

Arnold, L., 1974. Stochastic Differential Equations: Theory and Applications. Wiley, New York.

Albeverio, S., Klar, A., 1994. Long time behavior of nonlinear stochastic oscillators: the one dimensional Hamiltonian case. J. Math. Phys. 35, 4005–4027.

Albeverio, S., Kolokoltsov, V.N., 1997. The rate of escape for some Gaussian processes and scattering theory for their small perturbations. Stochastic Process. Appl. 67, 139–159.

Bernard, P., Wu, L., 1998. Stochastic linearization: the theory. J. Appl. Probab. 35, 718–730.

Baxter, J.R., Jain, N.C., Varadhan, S.R.S., 1991. Some familiar examples for which the large deviation principle does not hold. Comm. Pure Appl. Math. 34, 911–923.

Bryc, W., Smolenski, W., 1993. On the convergence of averages of mixing sequences. J. Theoret. Probab. 6, 473–483.

Bryc, W., Dembo, A., 1996. Large deviations and strong mixing. Ann. Inst. H. Poincaré Probab. Statist. 32, 549–569.

Dawson, D.W., Gartner, J., 1987. Long time fluctuation of weakly interacting diffusions. Stochastic 20, 247–308.

de Acosta, A., 1988. Large deviations for vector valued additive functionals of a Markov process: lower bound. Ann. Probab. 16, 925–960.

de Acosta, A., 1990. Large deviations for empirical measures of Markov chains. J. Theoret. Probab. 3, 395–431.

de Acosta, A., Chen, X., 1998. Moderate deviations for empirical measures of Markov chains. J. Theoret. Probab. 11 (4), 1075–1110.

Dembo, A., Zeitouni, O., 1998. Large deviations Techniques and Applications, 2nd Edition. Springer, New York.

Deuschel, J.D., Stroock, D.W., 1989. Large Deviations. Pure and Applied Mathematics, Vol. 137. Academic Press, New York.

Donsker, M.D., Varadhan, S.R.S., 1975. Asymptotic evaluation of certain Markov process expectations for large time, I-IV. Comm. Pure. Appl. Math. 28, 1–47, 279–301.

Donsker, M.D., Varadhan, S.R.S., 1976. Asymptotic evaluation of certain Markov process expectations for large time, I-IV. Comm. Pure. Appl. Math. 29, 389–461.

Donsker, M.D., Varadhan, S.R.S., 1983. Asymptotic evaluation of certain Markov process expectations for large time, I-IV. Comm. Pure. Appl. Math. 36, 183–212 (1983).

Down, D., Meyn, P., Tweedie, R., 1995. Exponential and uniform ergodicity of Markov processes. Ann. Probab. 23 (4), 1671–1691.

Ellis, R.S., 1985. Large Deviations and Statistical Mechanics. Springer, Berlin.

Freidlin, M., Weber, M., 1998. Random perturbation of nonlinear oscillators. Ann. Probab. 26 (3), 925–967.

Hilbert, A., 1990. Bounds on transition probabilities for a stochastically perturbed Hamiltonian system. In: Truman, A., Davies, I.M. (Eds.), Stochastic and Quantum Mechanics. World Scientific, Singapore, pp. 151–164.

Jacod, J., Shiryaev, A.N., 1987. Limit Theorems for Stochastic Processes. Grundlehren der mathematischen Wissenschaften, Vol. 288. Springer, Berlin.

Jain, N.C., 1990. Large deviations for additive functionals of Markov processes. Ann. Probab. 17 (3), 1073–1098.

Kato, T., 1984. Perturbation Theory For Linear Operators, 2nd Edition (2nd corrected printing). Springer, Berlin.

Khas'minskii, R.Z., 1980. Stochastic stability of Differential Equations. Sijthoff & Noordhoff, Alphen fa/d Rijn.

Meyn, P., Tweedie, R., 1993. Stability of Markovian processes II: continuous time processes and sampled chains. Adv. Appl. Probab. 25, 487–517.

Ney, P., Nummelin, E., 1987. Markov additive processes (I): eigenvalue properties and limit theorems; (II): large devaitions. Ann. Probab. 15, 561–592, 593–609.

Revuz, D., 1976. Markov Chains. North-Holland, Amsterdam.

Roberts, J.B., Spanos, P.D., 1990. Random Vibration and Statistical Linearization. Wiley, New York.

Wu, L., 1991. Grandes deviations pour les processus de Markov essentiellement irréductibles, I. temps discret. C.R. Acad. Sci. Sér. I 312, 608–614.

Wu, L., 1992. Grandes deviations pour les processus de Markov essentiellement irréductibles, II. temps continu. C.R. Acad. Sci. Sér. I 314, 941–946.

Wu, L., 1995. Moderate deviations of dependent random variables related to CLT. Ann. Probab. 23 (1), 420–445.

Wu, L., 2000. Uniformly integrable operators and large deviations for Markov processes. J. Funct. Anal. 172 (2), 301–376.